Mon. Not. R. Astron. Soc. **000**, 000–000 (0000) Printed 5 February 2008 (MN LATEX style file v2.2)

Selecting damped Lyman-α systems through Ca II absorption. I: Dust depletions and reddening at $z \sim 1$

Vivienne Wild[1,2⋆], Paul C. Hewett[1] and Max Pettini[1]

1. Institute of Astronomy, University of Cambridge, Cambridge CB3 0HA, UK
2. Max-Planck-Institut für Astrophysik, 85748 Garching, Germany

5 February 2008

ABSTRACT

We use the average *E(B−V)* and Zn II column densities of a sample of $z \sim 1$ Ca II $\lambda\lambda 3935, 3970$ absorption line systems selected from the fourth data release of the Sloan Digital Sky Survey (SDSS)to show that on average, with conservative assumptions regarding metallicities and dust-to-gas ratios, they contain column densities of neutral hydrogen greater than the damped Lyman-α (DLA) limit. We propose that selection by Ca II absorption is an effective way of identifying high column densities of neutral hydrogen, and thus large samples of DLAs at $z_{abs} \lesssim 1.3$ from the SDSS. The number density of strong Ca II absorbers (with rest-frame equivalent width $W_{\lambda 3935} \geqslant 0.5$ Å), is $\sim 20-30\%$ that of DLAs, after correcting for the significant bias against their detection due to obscuration of the background quasars by dust. On average these absorbers have $E(B-V) \gtrsim 0.1$ mag; the dustiest absorbers show depletions of refractory elements at a level of the largest depletions seen in DLAs. For the first time we can measure the dust-to-metals ratio in a sample of absorption selected galaxies, and find values close to, or even larger than, those observed locally. All of these properties suggest that a substantial fraction of the Ca II absorbers are more chemically evolved than typical DLAs. There is a trend of increasing dust content with $W_{\lambda 3935}$; this trend with strong-line equivalent width is also observed in an equivalent, but much larger, sample of Mg II absorbers. Such a trend would result if the dustier systems are hosted by more massive, or disturbed, galaxies. Follow-up imaging is required to provide conclusive evidence for or against these scenarios. From consideration of the *E(B−V)* distribution in our sample, and assuming Ca II absorbers represent a subset of DLAs, we calculate that dust obscuration causes an underestimation in the number density of DLAs by at least $8-12\%$ at these redshifts. Finally, the removal of Broad Absorption Line (BAL) quasars from the SDSS quasar sample increases the sensitivity of the detection of reddening by intervening absorbers. To this end, we describe a new, automated, principal component analysis (PCA) method for identifying BAL quasars.

Key words: dust, extinction - galaxies: ISM, abundances - quasars: absorption lines

1 INTRODUCTION

Measurements of the chemical composition of galaxies play an important role in understanding galaxy formation and evolution. The study of galaxy metallicities is closely intertwined with the question of their dust content: systems richer in metals have the greater potential to form dust grains. These grains selectively deplete metals in the interstellar medium (ISM) of the galaxies, redden their spectral energy distribution and cause an overall extinction of the light. The uncertainty in dust composition and distribution in galaxies affects the interpretation of many extragalactic observations.

Quasar absorption line systems provide a powerful probe of metals in the ISM of galaxies at high redshift, unhampered by luminosity bias, through the absorption of the light from a background quasar by gas phase metal ions. Notably, reddening due to dust has failed to be detected in samples of the highest column density absorption line systems, damped Lyman-α systems (DLAs), at $z \sim 2-4$ (Murphy & Liske 2004; Ellison et al. 2005). The relative paucity of dust particles in DLAs is backed up by observations of absorbers in radio selected quasar spectra, which indicate that extinction due to dust is a small effect (CORALS survey Ellison et al. 2001), and metal abundances show DLAs to be generally metal- and therefore dust-poor (e.g. Pettini 2004).

⋆ vwild@mpa-garching.mpg.de

1.1 Ca II in the local universe

This paper investigates the dust and metal properties of a new class of absorption line system, selected via the Ca II $\lambda\lambda 3935, 3970$ doublet[1] from the Sloan Digital Sky Survey (SDSS). These are the familiar K and H lines of the solar spectrum.

Ca is an α-capture element thought to be produced mainly by Type II supernovae. The convenient positioning of the resonance lines of its singly-ionised ion in the violet portion of the optical spectrum has made it accessible to stellar and interstellar studies for over a century, with major Galactic surveys by Adams (1949) and Marschall & Hobbs (1972). The inference of Ca abundance from Ca II absorption lines is however compromised by two considerations. First, with an ionisation potential of 11.9 eV, lower than that of H I, Ca^+ is a minor ionisation stage of Ca in the neutral ISM—most of the Ca is doubly ionised. Second, Ca is one of most depleted elements in the ISM—typically more than 99% of all Ca is 'hidden from view' having been incorporated into dust grains (Savage & Sembach 1996). The degree of depletion is expected to depend significantly on both the density of the gas (not just N(H I)) and on the presence of shocks, which may result in grain destruction, returning Ca to the gas phase. However, as stressed by Wakker & Mathis (2000), the dispersion in the measured gas-phase abundance of Ca II at fixed N(HI) in the Milky Way is remarkably small. Although the precise interpretation of the detection of a significant column density of Ca II is complex, it is clear that a large column density of neutral hydrogen is implied, given the high degree of depletion and low ionisation energy.

Ca II absorption in quasar spectra due to low redshift galaxies was detected early on in the history of quasar absorption line studies; for a review of early results see Blades (1988). The most comprehensive study was carried out by Bowen et al. (1991) who concluded that at projected separations $\gtrsim 20$ kpc from the centres of visible galaxies, Ca II absorption is patchy and, where it does occur, relatively weak, with the rest frame equivalent width of the strongest member of the Ca II doublet $W_{\lambda 3935} \lesssim 0.2$ Å. Examples of strong Ca II absorption ($W_{\lambda 3935} \gtrsim 0.5$ Å) were mostly confined to sightlines with projected separations $\lesssim 10$ kpc. Evidence for mergers was seen in those cases with large projected separations and large column densities of Ca II (Bowen 1991).

Since the early 1990s the focus of quasar absorption line research moved to high redshift and Ca II was largely ignored due to both its rarity and wavelength. At $z > 1$ the doublet moves into the near infrared which, at that time, was difficult to access. Recently, very strong Ca II absorption has been detected in a small number of highly reddened, $E(B-V) \sim 1$, sightlines to quasars (Petitjean et al. 2000; Hall et al. 2002; Wang et al. 2005). Such detections confirm the link between high gas column densities, dust and Ca II absorption in relatively extreme circumstances when the line-of-sight is almost coincident with the centre of a galaxy.

1.2 Metal absorbers and DLAs

DLAs are quasar absorption line systems with the highest column densities of neutral hydrogen, with a nominal column density limit of $N(\mathrm{H\,I}) > 2\times10^{20}\mathrm{cm}^{-2}$. Although they have been studied extensively since the seminal paper by Wolfe et al. (1986) twenty years ago, their precise nature, evolution and relation to the population of galaxies as a whole continue to be the subject of much discussion (Wolfe et al. 2005). Interest in DLAs results primarily from their apparent domination of the neutral gas mass fraction of the Universe at high redshift. Such cold, neutral gas is required for later star formation.

The defining diagnostic of a DLA, the Lyman-α absorption line, does not enter the optical atmospheric window until redshift $z_{\rm abs} \sim 1.8$. At $z \sim 0$ DLAs can be identified through 21 cm emission (e.g. Zwaan et al. 2005). The need for space-based observations has made it difficult to obtain large samples of intermediate redshift DLAs which are crucial both to clarify the relation of DLAs to luminous galaxies by direct imaging of the DLA hosts (which is much easier at redshifts $z \lesssim 1$), and to follow the evolution of their properties over the course of time.

Rao & Turnshek (2000) proposed a useful prognostic for intermediate redshift DLAs: the equivalent widths of the strong Mg II $\lambda 2796$ and Fe II $\lambda 2600$ absorption lines can be used to identify DLAs with a ∼40% success rate (Rao 2005). A complementary method to that proposed by Rao & Turnshek for recognising DLAs at intermediate redshift is through the detection of absorption lines which are intrinsically weak, due to low cosmic abundance, low transition probability, ionisation state or affinity for dust grains. By choosing lines appropriately, the success rate of identification can approach 100%, circumventing to some extent the need for UV spectroscopy; for example, the detection of the Zn II $\lambda\lambda 2026, 2062$ doublet virtually guarantees the DLA nature of an absorption system (Pettini et al. 1990).

As highlighted above, little is known of the strength of Ca II lines in DLAs; here we propose that approximately one fifth to one quarter of DLAs have Ca II lines with rest frame equivalent widths exceeding 0.5 Å for the stronger member of the doublet. The Sloan Digital Sky Survey (SDSS) quasar catalogue, with more than 45 000 spectra in its third data release (DR3, Schneider et al. 2005), provides the ideal database for obtaining large samples of metal line systems over a wide redshift range. The ability to detect significant numbers of DLAs using the SDSS over the entire redshift range $0 \lesssim z_{\rm abs} < 1.3$ would provide a much needed injection of new intermediate to low redshift DLAs for imaging and chemical evolution studies.

1.3 Aims of this paper

In a recent paper (Wild & Hewett 2005b, hereafter Paper I), we assembled a sample of 31 Ca II absorption systems selected from a high signal–to–noise ratio subset of the SDSS DR3 at redshifts $0.84 < z_{\rm abs} < 1.3$ and reported the detection of a significant degree of reddening—corresponding to a colour excess $E(B-V) = 0.06$—of the background quasars by dust in these absorbers. In this follow-up paper, we analyse the relative abundances of refractory elements in these intermediate redshift Ca II absorbers with a view to establishing their connection to the DLA population.

Specifically, we construct composite spectra which allow us to measure the relative abundances of Zn II, Cr II, Fe II, Mn II, Ti II and Ca II. The Zn measurement is crucial since, among the elements considered, it is the only one which shows little affinity for dust (Savage & Sembach 1996) and therefore gives a standard against which to compare the gas-phase abundances of the others. We add to the sample of Paper I six additional Ca II systems identified from the fourth data release (DR4) of the SDSS (Adelman-McCarthy et al. 2005), taking the total number of Ca II absorbers to 37. 27 of these have rest frame spectra that encompass the region of the Zn II doublet. We also re-examine in more detail the red-

[1] Vacuum wavelengths are used throughout this paper.

dening results of Paper I in light of the abundance determinations presented here.

The paper is organised as follows. In Section 2 we describe our search criteria for identifying absorption line systems. Strong Mg II $\lambda2796$ and Fe II $\lambda2600$ absorption lines are seen in all Ca II absorbers; we compare their equivalent width distributions with those of Mg II absorption systems at similar redshifts in SDSS quasars, paying particular consideration to the DLA selection criteria of Rao (2005). In Section 3 we calculate the redshift path of our Ca II survey. In Section 4 we derive column densities and relative abundances of elements detected in composite spectra of Ca II absorption line systems; we compare the abundance patterns with those seen in known DLAs and in the interstellar medium of the Milky Way. Section 5 deals with the reddening introduced by different samples of Ca II and Mg II absorbers on the observed spectral energy distributions (SEDs) of the background quasars. We discuss our results in Section 6, focusing in particular on the use of Ca II in the identification of DLAs, on the metallicities and dust content of Ca II-selected DLAs at $z \sim 1$, and on the implications of our findings for the more general issue of quasar obscuration by dust-rich intervening absorbers. Appendix A gives details of our new method for recognising quasars that show evidence of broad absorption line systems (BALs).

2 Ca II AND Mg II ABSORPTION SYSTEMS

The sample of quasar spectra used during the absorption line search and creation of control quasar composite spectra for the reddening analysis was restricted to those with spectroscopic signal–to–noise ratio (SNR) > 10 in the i-band and with Galactic extinction-corrected point spread function (PSF) magnitudes $i < 19.0$ in order to minimise the number of false detections during the Ca II line searches. The magnitude cut is very similar to that used in constructing the main spectroscopic quasar sample from the SDSS photometric catalogue ($i \leqslant 19.1$). After exclusion of BAL quasars, the final sample consists of 11 371 quasars from the DR3 quasar catalogue (Schneider et al. 2005) and a further 3 153 from the quasars observed on the 266 additional spectroscopic plates in DR4.

We searched the 14 524 quasars independently for both Ca II $\lambda\lambda3935, 3970$ and Mg II $\lambda\lambda2796, 2804$ absorption doublets at redshifts $0.84 < z_{\rm abs} < 1.3$. Residual sky OH emission features were subtracted from all spectra using the method of Wild & Hewett (2005a), which greatly reduces residual sky noise in the wavelength region around the Ca II lines. A "continuum" was defined for each quasar via the application of a simple 41–pixel median filter. The "difference" spectrum, to be searched for absorption features, was then obtained by subtracting the continuum from the original quasar spectrum. The absorption line search used a matched-filter technique (e.g. Hewett et al. 1985) with two template Gaussian doublets of the appropriate wavelength separation and full width at half maximum FWHM = 160 km s^{-1} (the resolution of the SDSS spectra), 200 km s^{-1} and 240 km s^{-1}. Three values of the doublet ratio were incorporated into the search: 2:1 (corresponding to unsaturated lines on the linear part of the curve of growth), 1:1 (for saturated lines on the flat part of the curve of growth) and an intermediate case, 4:3. At each pixel the template giving the largest SNR was determined and candidate absorbers were selected by applying a threshold value for the SNR (SNR$> 6\sigma$ for Mg II and SNR$> 5\sigma$ for Ca II).

The redshift range $0.84 < z_{\rm abs} < 1.3$ was determined in Paper I from consideration of, respectively, the lowest redshift at which the broad 2175 Å bump in the Milky Way reddening curve is detectable, and the highest redshift at which the Ca II doublet can be found in SDSS spectra. For the metal abundance analysis presented in this paper, we require coverage of the Zn II $\lambda\lambda2026, 2062$ lines associated with the Ca II systems; this increases the lower redshift limit to $z_{\rm abs} > 0.88$. We also exclude systems where the 2175 Å bump (or Zn II doublet) falls below the Lyman α emission line to avoid confusion with the Lyman α forest.

2.1 Ca II systems

At the redshifts of interest the Ca II lines fall in the red portion of the optical spectrum, beyond 7250 Å, where the SDSS quasar spectra become progressively noisier due to the increasing sky background and decreasing instrumental sensitivity. Thus, we required candidate Ca II systems to possess corresponding Mg II absorption within ±200 km s^{-1}. This list of Ca II candidates was subjected to further scrutiny by a fully parameterised fit of the absorption lines. A continuum was fitted to the regions around the Mg II and Ca II lines and the corresponding portions of the quasar spectra were normalised by dividing by the continuum level. Gaussian doublets were then fitted to the normalised spectra using a maximum-likelihood routine; the position and line width of the doublet, and the relative strengths of the individual lines, were allowed to vary freely. Rest frame equivalent widths (W) were calculated from the parameters of the Gaussian fits and errors estimated by propagation of the parameter errors derived during the maximum likelihood fit. For the purposes of composite construction (Sections 4.1 and 5.1) absorption redshifts were redefined using the centre of the fitted Mg II $\lambda2796$ line. A small number of Ca II doublets for which the fit proved to be visually unsatisfactory were removed from the sample.

Our final catalogue consists of 37 Ca II systems with $W_{\lambda3935} \gtrsim 0.35$ Å and $\langle z_{\rm abs} \rangle = 0.95$; 27 of these are suitable for use in the Zn II analysis, having $z_{\rm abs} > 0.88$. Fig. 1 shows some examples of spectra with Ca II systems in regions around the Ca II $\lambda\lambda3935, 3970$, Mg II $\lambda\lambda2796, 2804$ and Fe II $\lambda\lambda2587, 2600$ doublets. The rest frame equivalent widths of the Ca II and Mg II doublets, together with Fe II $\lambda2600$ and Mg I $\lambda2853$ for all the Ca II systems in the DR3 quasar catalogue can be found in Table 1 of Paper I. Table 1 of this paper gives the same information for the additional six Ca II systems found in DR4 quasar spectra.

2.2 Mg II systems

It is interesting to compare the properties of the rare Ca II absorbers with those of other classes of quasar absorption line systems. To this end, a catalogue of 2 338 Mg II absorbers was compiled from the sample used for confirmation of the DR3 Ca II absorption systems. The same fully parameterised fit was carried out as described above for the Ca II sample. However, because of the large number of systems, we automatically culled those absorbers from the sample with poor continuum fits, unphysical line ratios and detection significances (for the Mg II doublet) of less than 4, rather than relying on visual inspection of individual systems. Mg II absorbers also appearing in our Ca II catalogue were removed from this sample.

Among these Mg II systems we are especially interested in the subset satisfying the criteria for likely DLA candidates. Specifically, Rao (2005) found that 43% of strong ($W_{\lambda2796} > 0.6$Å) Mg II systems with associated Fe II $\lambda2600$ absorption such that

Table 1. Name and spectroscopic identification of each quasar in our DR4 Ca II absorber sample, together with rest frame equivalent widths of Ca II $\lambda\lambda 3935, 3970$, Mg II $\lambda\lambda 2796, 2804$, Mg I $\lambda 2853$ and Fe II $\lambda 2600$. The final column gives the estimated reddening of each quasar by the Ca II absorber, calculated assuming an extinction curve similar to that of the Large Magellanic Cloud (see Section 5.3). See Table 1 of Paper I for equivalent details of the DR3 Ca II systems.

SDSS ID	MJD[a],plate,fibre	i^b	$z_{\rm em}$	$z_{\rm abs}{}^c$	$W_{\lambda 3935,3970}$	err(W)	$W_{\lambda 2796,2803}$	$W_{\lambda 2853}$	$W_{\lambda 2600}$	$E(B-V)$
J080958.56+515118.0	53090,1780,532	18.56	1.289	0.902	0.64, 0.39	0.13, 0.11	2.25, 2.22	1.14	1.92	−0.007
J120300.96+063440.8	53089,1623,209	18.47	2.182	0.862	1.42, 0.89	0.29, 0.24	5.57, 4.97	2.98	3.75	0.417
J122756.40+425631.2	53112,1452,505	17.17	1.310	1.045	0.39, 0.17	0.07, 0.08	1.61, 1.39	0.25	1.15	0.026
J130841.28+133130.0	53089,1772,122	18.60	1.954	0.951	0.87, 0.91	0.27, 0.29	1.82, 1.61	0.68	1.16	0.080
J140444.16+551637.2	53088,1324,421	18.46	1.589	1.070	1.33, 0.54	0.34, 0.22	1.98, 2.11	0.94	1.44	0.195
J153503.36+311832.4	53119,1388,068	17.79	1.510	0.904	0.84, 0.36	0.15, 0.12	2.11, 1.80	0.82	1.03	−0.008

[a] Modified Julian date
[b] PSF i-band magnitude, corrected for Galactic extinction
[c] Measured from Mg II $\lambda 2796$

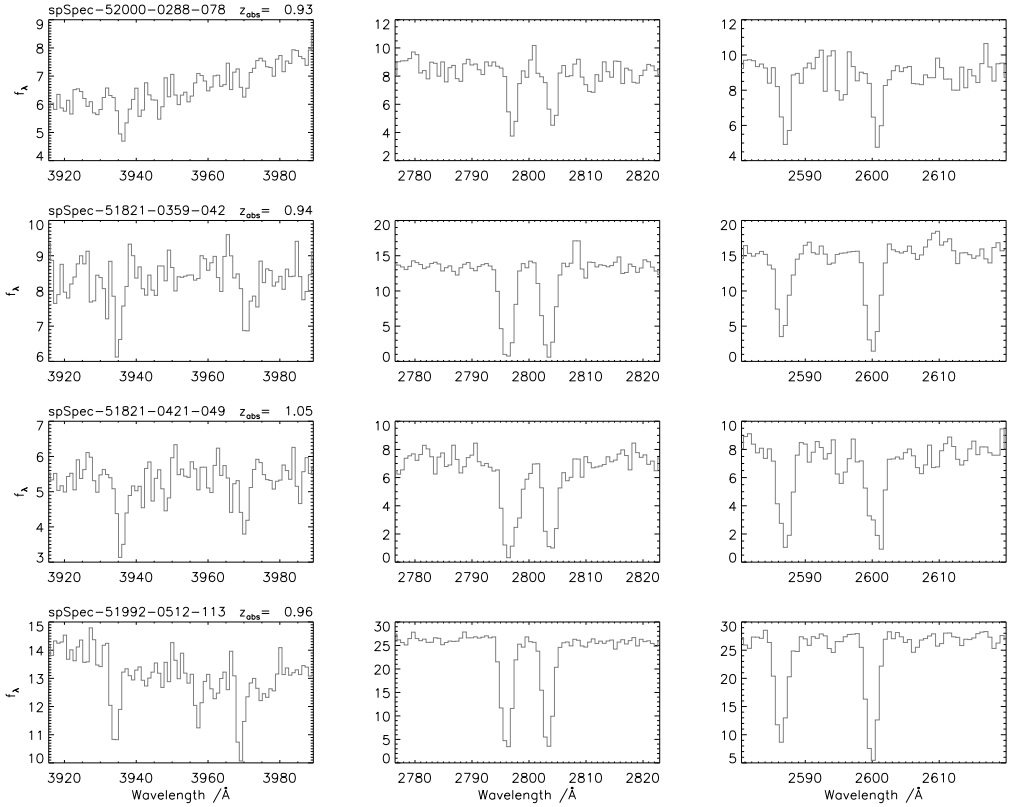

Figure 1. Examples of Ca II systems (left panels) found in the SDSS quasar sample, along with the associated Mg II (centre) and Fe II lines (right). The units of the y-axis are 10^{-17} erg cm^{-2} s^{-1} Å^{-1}. The SDSS spectroscopic filename (MJD, plate number, fibre number) and absorber redshift are given above each panel in the left-hand column.

$1 < W_{\lambda 2796}/W_{\lambda 2600} < 2$ are confirmed to be *bona-fide* DLAs by subsequent ultraviolet (UV) spectroscopy. 789 of our Mg II systems with $0.84 < z_{\rm abs} < 1.3$ satisfy these requirements and we shall refer to this sample as the "Mg II-selected DLAs".

2.3 Strong metal lines associated with Ca II absorption systems

In Fig. 2, the equivalent widths of Mg II $\lambda 2796$, Fe II $\lambda 2600$ and Mg I $\lambda 2853$ in the Ca II absorbers (filled symbols) are compared with those of the Mg II absorbers (small dots). In general, Ca II systems tend to have strong associated Mg II, Fe II and Mg I lines. Referring to the left-hand panel, it can seen that all but two of the Ca II systems fall within the defining criteria of Mg II-selected DLAs and their average equivalent width is significantly higher

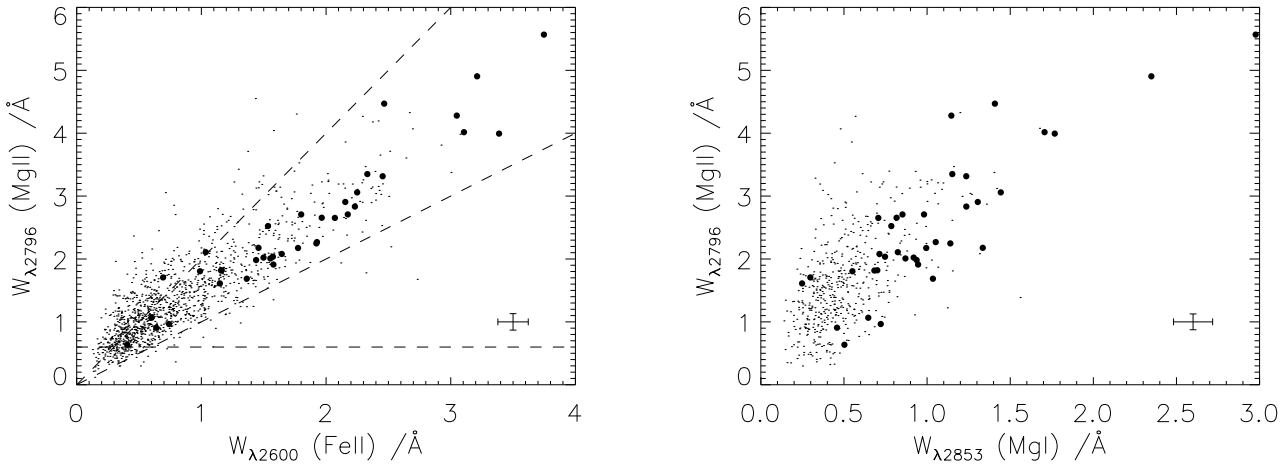

Figure 2. The rest frame equivalent widths of Mg II λ2796 vs. Fe II λ2600 on the left and Mg I λ2853 on the right for the Ca II absorbers (filled symbols) and Mg II absorbers (small points). The dashed lines in the left panel indicate the limits within which DLAs have been found in Mg II absorbers by Rao (2005). The error bars in the bottom right-hand corners show the typical errors on the equivalent widths.

than the average for the Mg II-selected DLAs in general. Turning to the right-hand panel, there is a hint that the Ca II absorbers may have stronger Mg I for a given equivalent width of Mg II than conventional Mg II absorbers, possibly indicating that they tend to arise in regions of high gas density (Hobbs 1974, see also Section 6.4.1).

3 SELECTION FUNCTION AND TOTAL REDSHIFT PATH

In order to assess the statistical properties of our Ca II absorption line systems and compare them with those of other classes of absorbers, it is necessary to quantify the sensitivity of our survey as a function of Ca II equivalent width and wavelength. We have addressed this question with Monte Carlo simulations by placing artificial Ca II and Mg II absorption lines in the SDSS quasar spectra and determining the fraction recovered using the same search techniques as in the real survey. The simulations were run for doublets with ratios of 1:1, 2:1 and 4:3 for Ca II, and 1:1 and 2:1 for Mg II and the effect of varying the line width was investigated. The general conclusion from these different trials was that for the parameter ranges considered there was a variation in detection efficiency of at most 5-10%.

Fig. 3 shows the probability of detection, $P_{\rm Ca}$, of a Ca II system with doublet ratio $W_{\lambda3935}$: $W_{\lambda3970}$ =4:3 as a function of redshift $z_{\rm abs}$ and equivalent width of the stronger member of the doublet, $W_{\lambda3935}$. In this example we have assumed that a Mg II doublet is detected for every artificial Ca II system recovered (thus setting $P_{\rm Mg} = 1$), except at redshifts which place the Mg II lines close to strong sky emission lines. Values of $z_{\rm abs}$ and $W_{\lambda3935}$ for the 37 Ca II systems are overplotted as filled circles. The probability of detection drops sharply at redshifts $z_{\rm abs} > 1.2$, as the Ca II doublet moves beyond 8500 Å in the observed frame where the spectra are of poorer quality; the pathlength available for finding absorbers also decreases with increasing redshift. Even so, it is perhaps surprising that *no* Ca II systems have been detected at these redshifts. We test the significance of this result below.

By multiplying the number of quasar spectra in which a Ca II absorption system at a given redshift could be detected, $\mathcal{N}_{\rm quasar}$ $(z_{\rm abs})$, by the probability of detection given its values of

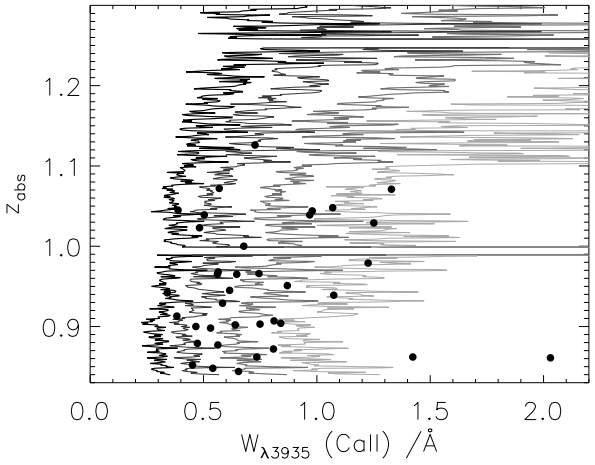

Figure 3. The probability of detecting a Ca II absorption system in our survey as a function of absorption redshift and rest equivalent width of the stronger member of the Ca II doublet, calculated from a series of Monte Carlo trials. From left to right, the contours are drawn at probabilities $P_{\rm Ca} = 0.1, 0.3, 0.7$ and 0.95. In the example shown here, we assumed a Ca II doublet ratio $W_{\lambda3935}$: $W_{\lambda3970} = 4 : 3$ and that a Mg II doublet is detected for each Ca II system recovered by our search technique, except when blended with the strong sky lines at λ5579 and λ6302. These sky lines give rise to the conspicuous gaps in our detection efficiency at $z_{\rm abs} \simeq 0.99$ and 1.25. Filled circles correspond to the 37 Ca II absorbers detected in our survey.

$z_{\rm abs}$, $W_{\lambda3935}$ and $W_{\lambda2796}$, we obtain an estimate of the *effective* number of sightlines over which we could expect to find such a system:

$$\mathcal{N}_{\rm eff}(z_{\rm abs}, W_{Ca}, W_{Mg}) = \mathcal{N}_{\rm quasar}(z_{\rm abs}) \times P_{\rm detect} \quad (1)$$

where

$$P_{\rm detect} = P_{\rm Ca}(W_{\rm Ca}, z_{\rm abs}) \times P_{\rm Mg}(W_{\rm Mg}, z_{\rm abs}) \quad (2)$$

and the notation for the rest frame equivalent widths of Ca II λ3935 and Mg II λ2796 has been shortened for convenience.

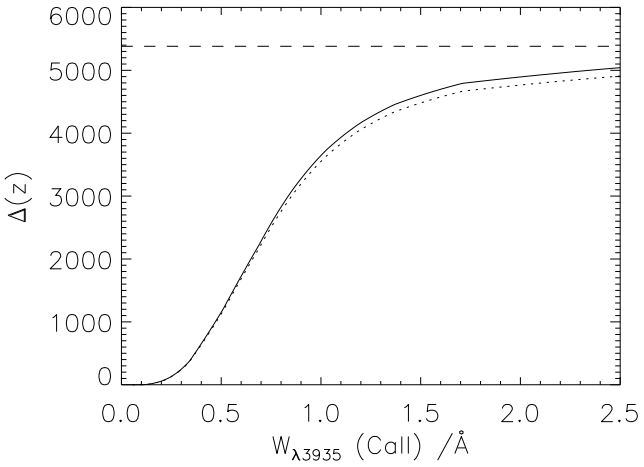

Figure 4. The total redshift path of our Ca II survey calculated from Monte Carlo trials. In the two examples shown here, the artificial Ca II lines used in the simulations had a fixed doublet ratio of 4:3, and the corresponding Mg II lines were either always detected (continuous curve) or assumed to have $W_{\lambda 2796} = W_{\lambda 2803} = 1.0$ Å (dotted curve). The horizontal long-dash line shows the maximum pathlength available if the probability of detection were $P_{\rm detect} = 1$ over the entire redshift path covered by each input quasar spectrum.

Integrating over redshift gives the total redshift path covered by our survey as a function of rest frame equivalent width:

$$\Delta Z(W_{\rm Ca}, W_{\rm Mg}) = \int_{z_{\rm abs,min}}^{z_{\rm abs,max}} \mathcal{N}_{\rm eff}\,(z_{\rm abs}, W_{\rm Ca}, W_{\rm Mg})\,d\,z_{\rm abs}. \quad (3)$$

This function is plotted in Fig. 4 for two cases: one calculated assuming $P_{\rm Mg} = 1$ throughout (except when the Mg II lines are blended with sky lines) and the other adopting $W_{\lambda 2796} = 1.0$ Å (a conservative lower limit for the majority of Ca II absorbers) and a Mg II doublet ratio of 1:1. The difference between the two cases is minimal.

It was noted above that no absorption systems are found between $1.2 < z_{\rm abs} < 1.3$ and we are now in a position to calculate the significance of this result. The total pathlength available for finding absorbers in this redshift range is only 0.064 of that available over the entire $0.84 < z_{\rm abs} < 1.3$ range. By adopting a binomial distribution with 37 trials and P($z_{\rm abs} > 1.2$) = 0.064, the probability of detecting 0/37 absorbers in this redshift range is 0.086, i.e. the significance of the lack of detection is $\sim 90\%$. Perhaps with future SDSS catalogue releases it will be possible to confirm that the effect is indeed a low significance statistical fluctuation.

By combining the observed equivalent width distribution with Eq. 3, an estimate can be made of the intrinsic equivalent width distribution of the Ca II absorbers. This is shown alongside the Mg II equivalent width distribution in the two top panels of Fig. 5. We only show the completeness corrections above 0.5 Å and 0.3 Å for the Ca II and Mg II samples respectively; below these values the corrections become large and uncertain due to the small number of systems.

Finally, a statistic commonly used in absorption line surveys is the number of systems per unit redshift with rest equivalent width greater than some threshold value: $n(W^{\rm lim})$. While in general $n(W^{\rm lim})$ is a function of redshift, we assume negligible redshift dependence here since we are dealing with only a small redshift interval. By setting the minimum equivalent width of Mg II $\lambda 2796$ to be 0.6 Å, smaller than measured for any of the Ca II systems in our survey, we can remove the joint dependency in the cumulative probability function:

$$n(W_{\rm Ca}^{\rm lim}, W_{\rm Mg}^{\rm lim} = 0.6) = \sum_i \frac{1}{\Delta Z_i(W_{\rm Ca} \geqslant W_{\rm Ca}^{\rm lim})} \quad (4)$$

where the sum is over all absorbers in the sample with $W_{\lambda 3935}$ greater than the limit. The function $n(W_{\rm Ca}^{\rm lim}, W_{\rm Mg}^{\rm lim} = 0.6)$ is shown in the bottom left-hand panel of Fig. 5; we find that for $W_{\rm Ca\,II\,\lambda 3935}^{\rm lim} = 0.5$ Å, the Ca II systems have a number density per unit redshift $n(z) = 0.013$ (at a mean redshift $\langle z_{\rm abs} \rangle = 0.95$).

We also calculated $n(W^{\rm lim})$ for the Mg II systems (Section 2.2); the results are shown in the bottom right-hand panel of Fig. 5. Our estimates are in good agreement with those determined by Nestor et al. (2005) from the SDSS EDR (dashed line) over a very similar redshift range.

4 ELEMENT RATIOS IN Ca II ABSORBERS

4.1 Composite spectra

In this section we deduce the column densities and relative abundances of different ions associated with the Ca II absorbers whose absorption lines are covered in the SDSS spectra. The chemical composition of the Ca II systems should help clarify their relationship to other classes of quasar absorption line systems, particularly DLAs. We are especially interested in investigating the depletion pattern of refractory elements for comparison with the reddening results reported in Paper I and re-examined in Section 5 below. The unsaturated absorption lines which are required for precise abundance analyses are generally too weak to be detected and measured in the individual SDSS quasar spectra at our disposal. We circumvented this difficulty by adding together, in the absorber rest frame, the spectra of quasars with Ca II systems to form composites of sufficient SNR. This technique has also been applied to absorption systems in SDSS spectra by Nestor et al. (2003).

Out of the 37 quasar spectra in our Ca II catalogue, 27 encompass the redshifted wavelengths of the Zn II $\lambda\lambda 2026, 2062$ doublet in the Ca II absorption system. Given the diagnostic importance of this ion (e.g Pettini et al. 1990), we limited the abundance analysis to this subset of 27 Ca II absorbers. We considered three composite spectra: one containing all 27 absorbers, and two further subsets each containing about half of the spectra (13 or 14) separated at the median value of the equivalent width of the Ca II $\lambda 3935$ line, $W_{\lambda 3935} = 0.68$ Å. We refer to these three samples respectively as 'All', 'High-$W_{\lambda 3935}$', and 'Low-$W_{\lambda 3935}$'.

To create the composite spectra, the individual quasar spectra were shifted to the absorber rest frame without rebinning[2]. Large scale variations in each spectrum were removed by employing first a spline continuum fit and then a sliding median filter of 45 pixels to the residual, with strong absorption lines masked. The spectra were subsequently combined into a composite using an error-weighted arithmetic mean. Alternative methods for both flattening the individual spectra and combining them were explored; in all cases we found the final result to be insensitive to the precise methods used. A second normalisation was carried out on portions of the composite spectra that included absorption lines of interest by spline-fitting

[2] The SDSS spectra are binned in $\log \lambda$ with a pixel size of 0.0001, thus on blueshifting, rebinning of each spectrum is not necessary.

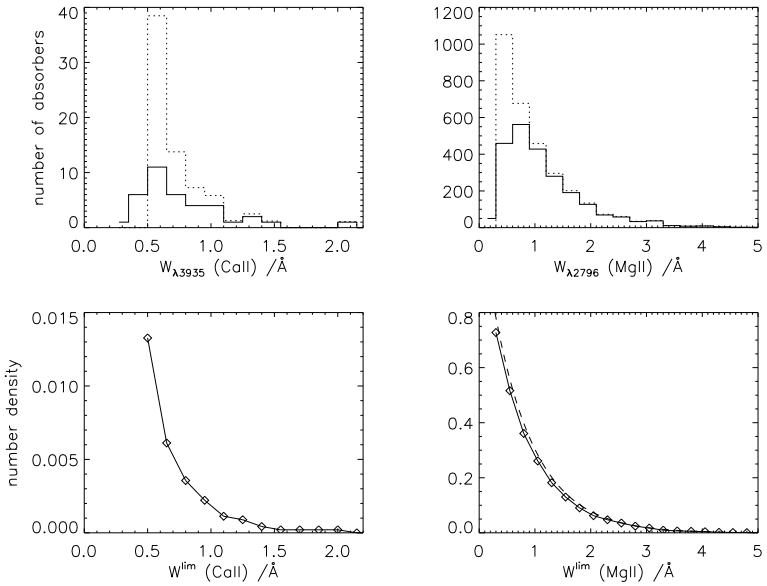

Figure 5. *Top two panels*: Equivalent width distributions of Ca II and Mg II absorption systems before (continuous histograms) and after (dashed histograms) completeness corrections. No corrections are shown for the smallest equivalent width bins where the statistics are too poor for reliable estimates. *Bottom two panels*: Number density of absorbers per unit redshift interval as a function of minimum equivalent width limit,$n(W^{\rm lim})$. The values we deduce for the Mg II systems are in good agreement with the fit to the distribution of SDSS EDR Mg II systems with $0.871 < z_{\rm abs} < 1.311$ (dashed line) reported by Nestor et al. (2005).

of the continuum, and the rest frame equivalent widths of the lines were then measured. Table 2 summarises the results, and Fig. 6 shows the normalised composite spectrum formed by coadding all 27 Ca II absorbers.

4.2 Element abundances and dust depletions

Inspection of Table 2 and Fig. 6 shows that our composite spectra cover a variety of transitions from the elements Mg, Ca, Ti, Cr, Mn, Fe and Zn. To avoid potential problems with line saturation, we limited our analysis to lines with apparent optical depth less than ~ 0.1, which are most likely to fall on the linear part of the curve of growth. Unless the distribution of equivalent widths of these lines among the 27 absorption systems is highly non-uniform—something which we can exclude from inspection of the individual spectra—the column densities we deduce from these weak transitions should be representative of the typical values for the sample as a whole. In any case, any non-linearities which may be affecting the absorption lines in the composite spectra would lead us to *under*estimate the corresponding column densities.

Ion column densities, N, were derived by fitting all the weak lines in the composite spectra with the line profile fitting package VPFIT[3] adopting f-values and rest wavelengths from the compilation by Morton (2003) and reproduced in Table 2. The profile fitting approach was chosen primarily to ensure internal consistency in absorption redshift and width among the lines; for such weak lines the column density is independent of the details of the line profile and essentially the same values of N would have been deduced with the analytical expression which relates equivalent width and column density for lines on the linear part of the curve of growth:

$$N = 1.13 \times 10^{20} \cdot \frac{W}{\lambda^2 f} \ \ {\rm cm}^{-2} \qquad (5)$$

where W and λ are both in Ångstroms.

Columns 3-5 of Table 3 list ion column densities in each of the three composites together with the errors returned by VPFIT. In all but one case (Ca II), the ions observed are the major ionisation stages of the corresponding elements in H I regions. Thus, under the working hypothesis that Ca II absorbers with equivalent width $W_{\lambda3935} \gtrsim 0.35$ Å (the lower limit of our sample) are DLAs (where the gas is mainly neutral and ionisation corrections are small), we take the ion ratios relative to Zn II as measures of the elements' abundances relative to Zn. Comparison with the solar scale [we adopt the set of values proposed by Lodders (2003)] finally gives the entries of [X/Zn][4] in the last three columns of Table 3.

While we cannot deduce the overall metal abundances of the Ca II absorbers without an independent measure of N(H I), we can still compare their depletion pattern to that commonly encountered in DLAs. We have referred the ratios to Zn II column density because, alone among the elements available to us, the depletion level of Zn onto dust grains is low in the Milky Way ISM, compared to the large and variable depletions of the refractory Ti, Cr, Mn and Fe relative to the solar values (Savage & Sembach 1996). In reality, intrinsic differences in the nucleosynthetic history of the elements can lead to departures from solar relative abundances; dust depletion subsequently operates on this underlying pattern. While this is a rich topic of study (e.g. Wolfe et al. 2005), it is not necessary to differentiate between the two effects in the current study as we are primarily concerned with an overall comparison between the Ca II absorbers and DLAs.

[3] http://www.ast.cam.ac.uk/∼rfc/vpfit.html

[4] We use the usual notation whereby $[{\rm X/Zn}] \equiv \log({\rm X/Zn})_{\rm obs} - \log({\rm X/Zn})_{\odot}$.

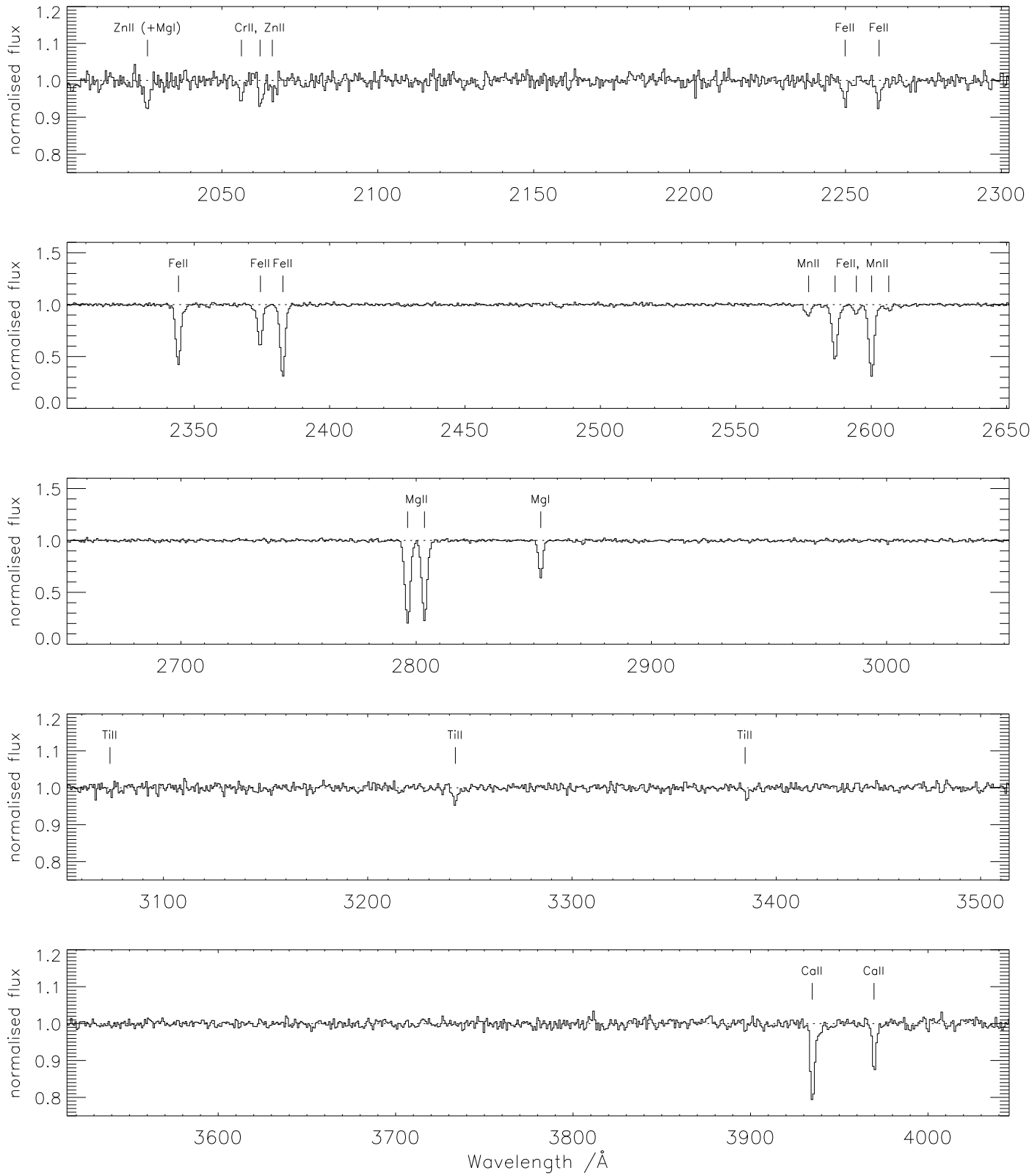

Figure 6. Composite spectrum of the 27 Ca II absorption systems with $\langle z_{\rm abs} \rangle = 0.98$; transitions listed in Table 2 are indicated. Note that two different y-axis scalings are used for the weak and strong lines respectively.

Table 2. Rest frame equivalent widths of metal lines measured in the three Ca II absorber composites.

			W (mÅ)		
Ion	Wavelength (Å)	f-value[a]	All	High-$W_{\lambda3935}$	Low-$W_{\lambda3935}$
Ca II	3934.775	0.6267	725± 30	1011± 45	489± 24
Ca II	3969.590	0.3116	352± 29	549± 44	371± 29
Ti II	3242.918	0.232	124± 14	120± 22	92± 19
Ti II	3384.730	0.358	76± 15	...	87± 18
Mg I	2852.963	1.83	782± 13	928± 30	717± 20
Mg II	2796.354	0.615	2256± 13	2472± 28	2191± 22
Mg II	2803.531	0.306	2106± 13	2238± 28	2067± 22
Fe II	2600.173	0.239	1722± 15	1799± 27	1598± 18
Fe II	2586.650	0.0691	1357± 17	1351± 26	1268± 19
Mn II	2606.462	0.198	147± 12	122± 21	100± 17
Mn II	2594.499	0.280	213± 14	231± 21	146± 17
Mn II	2576.877	0.361	277± 14	395± 29	227± 17
Fe II	2382.765	0.320	1582± 16	1606± 28	1554± 19
Fe II	2374.461	0.0313	929± 16	1014± 30	891± 20
Fe II	2344.214	0.114	1325± 16	1360± 29	1385± 24
Fe II	2260.781	0.00244	128± 17	129± 30	135± 20
Fe II	2249.877	0.00182	109± 15	95± 27	87± 17
Cr II	2066.164	0.0512	46± 12	80± 16	84± 17
Cr II (+Zn II)	2062.236	0.0759 (0.246)	114± 15	195± 22	111± 17
Cr II	2056.257	0.103	92± 18	90± 25	108± 19
Zn II (+Mg I)	2026.137	0.501 (0.113)	174± 19	267± 26	117± 19

[a] Rest frame wavelengths and f-values from Morton (2003)

Table 3. Ion column densities and element abundances relative to Zn.

		$\log N$			[X/Zn]		
Ion	Solar[a]	All	High-$W_{\lambda3935}$	Low-$W_{\lambda3935}$	All	High-$W_{\lambda3935}$	Low-$W_{\lambda3935}$
Zn II	4.63	12.82±0.065	13.16±0.054	12.59±0.110	0.00	0.00	0.00
Cr II	5.65	13.39±0.068	13.49±0.093	13.42±0.065	−0.45	−0.69	−0.19
Fe II	7.47	15.09±0.045	15.02±0.086	15.09±0.048	−0.56	−0.98	−0.34
Ti II	4.92	12.48±0.083	12.31±0.123	12.45±0.069	−0.63	−1.14	−0.43
Ca II	6.34	12.94±0.017	13.10±0.019	12.86±0.023	> -1.59 [b]	> -1.77	> -1.45
Mn II	5.50	13.09±0.023	13.19±0.030	12.96±0.032	−0.60	−0.84	−0.50

[a] Solar abundance relative to hydrogen, in the usual logarithmic scale with H at 12.00 (Lodders 2003).
[b] The values for Ca II are lower limits because a significant but unknown fraction of Ca is doubly ionised and the line ratios of some individual absorbers suggest a degree of saturation.

To this end, we have searched the available DLA literature for systems where the abundances of Zn and at least one of the four refractory elements considered here have been measured. After adjusting the abundances to Lodder's (2003) solar scale, we compare them in Fig. 7 with those of the Ca II absorbers. References to the original papers are given in the figure caption. Note that in general the measurements from the literature refer to a much wider range of redshifts than probed here.

Fig. 7 shows clearly that the element depletions deduced for the Ca II systems are typical of the values encountered in most DLAs, with Ti, Cr, Mn and Fe on average less abundant than Zn by factors of 3-4 (~ 0.5-0.6 dex). Their is a marked distinction between the two subsamples, with the 'High-$W_{\lambda3935}$' subsample in particular exhibiting some of the most pronounced depletions measured in DLAs up to now. It is certainly of considerable interest that similar levels of [Fe/Zn] depletion are seen in DLAs with significant molecular hydrogen contents, where dust is believed to be the catalyst for molecular hydrogen formation (Ledoux et al. 2003). Most data are available for the Zn and Cr pair (top panel of Fig. 7), partly for historical reasons (Pettini et al. 1990) and partly because, at the redshifts of most DLAs, the relevant absorption lines are conveniently located in the optical spectrum. The ratio [Cr/Zn] shows an approximate correlation with [Zn/H]: the depletion of Cr decreases with decreasing overall metallicity, although the scatter is considerable (Pettini et al. 1997; Akerman et al. 2005). By using this correlation as a rough indicator of metallicity for the Ca II absorbers, a range between $\sim 1/30$ and $\sim 1/3$ of solar is suggested ([Zn/H] between -1.5 and -0.5).

Fig. 8 presents the depletions relative to Zn of the Ca II absorbers compared to typical diffuse clouds in the solar neighbour-

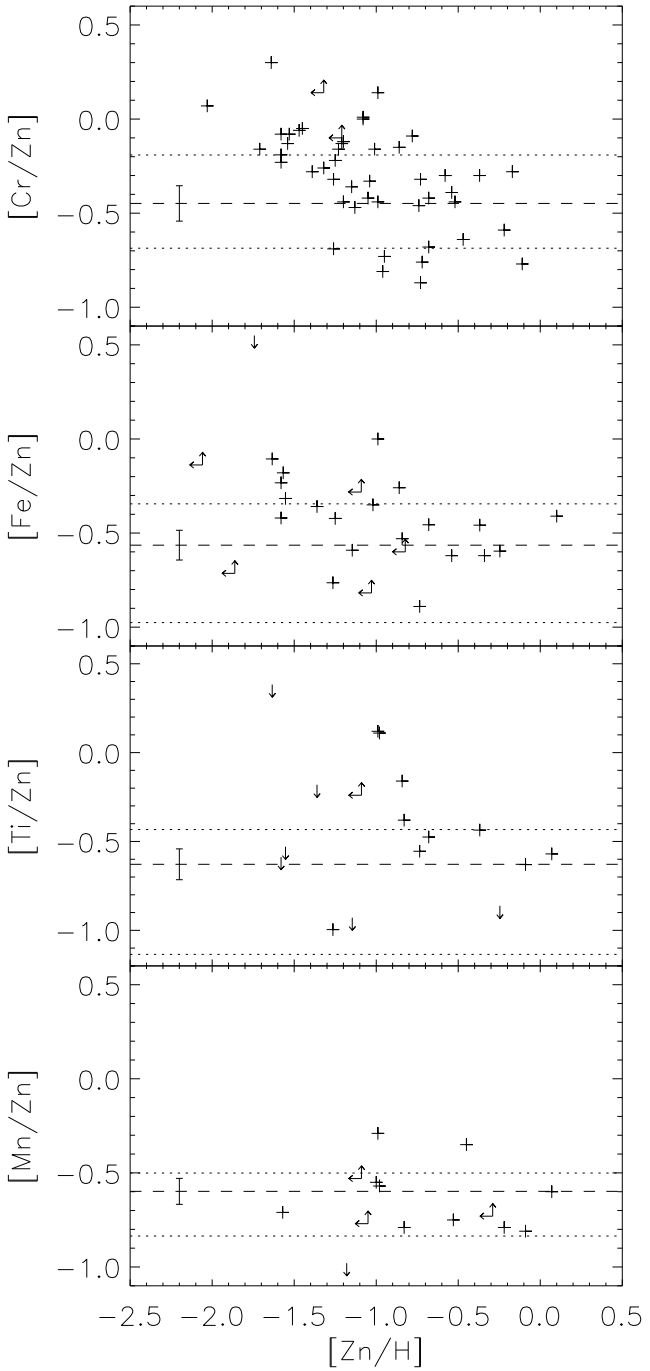

Figure 7. The crosses show the abundances of refractory elements relative to Zn (an element which shows little affinity for dust) in DLAs from measurements reported in the literature. The horizontal lines are plotted at the values of [X/Zn] in Table 3: the dashed lines (with error bars) are the values for the sample of all 27 Ca II absorbers, while the dotted lines are for the two subsamples of 'High-' (lower dotted line) and 'Low- $W_{\lambda 3935}$' (upper dotted line). The stronger Ca II absorbers consistently exhibit a higher degree of dust depletion. The DLA measurements generally refer to a wider redshift range than that of the Ca II sample considered here. The sources of the DLA data and corresponding redshift intervals are as follows. Cr: Kulkarni et al. (2005); Akerman et al. (2005) ($0.69 < z_{\rm abs} < 3.39$). Fe: Prochaska et al. (2001); Pettini et al. (1999); Pettini et al. (2000) ($0.61 < z_{\rm abs} < 3.39$). Ti: Ledoux et al. (2002); Prochaska et al. (2001) ($0.43 < z_{\rm abs} < 2.48$). Mn: Ledoux et al. (2002) ($0.43 < z_{\rm abs} < 2.14$). Many of these measurements were obtained from the HIRES DLA database at: http://kingpin.ucsd.edu/∼hiresdla/ .

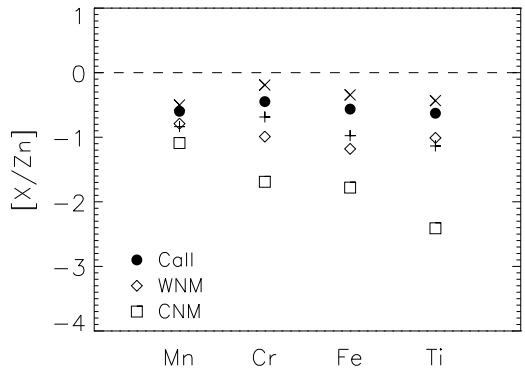

Figure 8. Abundances relative to Zn of refractory elements in the Ca II absorbers (indicated by crosses), and in the warm (diamonds) and cold (squares) neutral interstellar medium of the Milky Way (from the compilation by Welty et al. 1999). Three sets of crosses are shown, respectively for all Ca II systems (black/middle crosses), and for the 'Low-' (red/upper) and 'High-$W_{\lambda 3935}$' (blue/lower) subsamples. Elements are ordered by increasing condensation temperature (Savage & Sembach 1996) and corrected to the solar reference values used in this paper.

hood. The overall level of depletion is lower in the Ca II absorbers, similar to results for DLAs which is cited as being consistent with their lower metallicities (Vladilo 2004). However, we note that the depletions deduced for the 'High-$W_{\lambda 3935}$' subsample approach or exceed those typical of the warm neutral medium of the Milky Way.

5 REDDENING DUE TO DUST

An alternative method of estimating the dust content of quasar absorbers is to consider the reddening effect of the dust on the SED of the background quasar. This is an intrinsically statistical approach, because we do not have an *a priori* knowledge of the unobscured SED of an individual quasar, but rather we have to rely on the typical SED of an ensemble of quasars with properties broadly similar to those being tested for extinction. For this reason, the availability of large control samples of quasar spectra afforded by the SDSS has allowed such tests to be carried out to unprecedented levels of sensitivity. Murphy & Liske (2004) concluded that high redshift ($z_{\rm abs} \sim 3$) DLAs in the SDSS DR2 have an average *E(B−V)*$\lesssim 0.02$ based on the lack of any reddening signal. On the other hand, in Paper I we reported an average *E(B−V)*$= 0.06$ due to the 31 Ca II absorbers at $\langle z_{\rm abs} \rangle = 0.95$ in SDSS DR3. We also found that splitting that Ca II sample at $W_{\lambda 3935} = 0.7$ Å resulted in *E(B−V)*$= 0.099$ and 0.025 for the high and low equivalent width subsets respectively—a trend which we now see to be consistent with that of the element depletions discussed above.

In this paper we extend the analysis of Paper I in two ways. First, we perform our reddening analysis on the expanded sample of 37 Ca II absorbers, which includes the six newly discovered systems from DR4, and on the 27 Ca II systems analysed in the previous section to allow direct comparisons of *E(B−V)* values with Zn column densities and depletion patterns. Second, we repeat the analysis on the Mg II-selected DLAs [those Mg II absorption line systems fulfilling the criteria of Rao (2005), Section 2.2] for comparison with the Ca II absorbers and with the confirmed DLAs at $z \sim 3$ considered by Murphy & Liske (2004).

5.1 Quasar reference spectra and measurement of average reddening

We look for a reddening signal in the spectra of quasars with an intervening absorber by comparing their SEDs with that of a control spectrum: a composite constructed by averaging many spectra of quasars without those absorbers. The large number of quasars in the SDSS allows us to create control composites in small steps of quasar emission redshift $z_{\rm em}$. Details of the method have been described in Paper I and are only summarised here. Using the same sample of SDSS DR3 quasars which were searched for Ca II absorption systems (Section 2), but removing all spectra with identified Ca II or Mg II absorption lines, we constructed composite spectra in bins of $\Delta z_{\rm em} = 0.1$ staggered by $z = 0.05$. Each spectrum was first corrected for Galactic reddening using the quoted extinction in the SDSS photometric catalogue and the Galactic extinction curve from Cardelli et al. (1989) as extended by O'Donnell (1994). The spectra were then shifted to the quasar rest frame without rebinning, being careful to allow for the flux/Å term in the SDSS spectra, scaled by the median flux over a common rest wavelength range for the particular redshift bin (avoiding quasar emission lines, strong sky lines and bad pixels), and finally combined into a composite spectrum using a simple arithmetic mean. Different methods of combining spectra were tested, and we concluded that the details of the procedure made no significant difference to the final composites.

The procedure adopted here differs from that of Paper I in two small details. First, when constructing the reference spectra we excluded all quasars with Mg II absorbers (as well as the few quasars with Ca II systems), since we will be testing for reddening caused by both classes of absorption line system. Second, we dropped the correction for potential dependence of the quasar SED on apparent magnitude, since any effect was found to be insignificant.

Turning now to the spectra of quasars *with* absorbers: each one of these was shifted to the rest frame of the quasar, scaled by the median flux as above, and then divided by the reference composite quasar spectrum closest in redshift. The resulting "difference" spectrum, which contains any reddening signal, was then shifted to the rest frame of the absorber, $z_{\rm abs}$, before being combined with others to form the composite spectra to be analysed for reddening. This combination was performed using an arithmetic mean; a popular alternative for this type of analysis is the geometric mean. Employing a geometric rather than arithmetic mean results in reddening values 10-20% higher due to enhanced weighting given to the difference spectra with the largest slopes towards the blue end. In the composites studied in this paper the differences are insignificant, in almost all cases they are within the 1σ errors (68th percentile, see 5.2).

We constructed several different composite spectra for both the Ca II absorbers and Mg II-selected DLAs. For the former, we produced three composites equivalent to the 'All', 'High-$W_{\lambda 3935}$', and 'Low-$W_{\lambda 3935}$' samples of the previous section but containing all 37 absorbers. We also produced reddening composites for the three samples containing absorbers with available Zn II for direct comparison with the measured element column densities. For the Mg II-selected DLAs, we considered the full sample of 789 quasars and three subsamples with increasing absorption line strength. The subsamples were defined by drawing diagonal lines across Fig. 2 from $(W_{\lambda 2796},\, W_{\lambda 2600}) = (m, 0)$ to $(W_{\lambda 2796},\, W_{\lambda 2600}) = (0, m)$, where $m = 1.5, 2$ and 3 Å, and taking only the systems that lie above each line and are within the Rao (2005) limits.

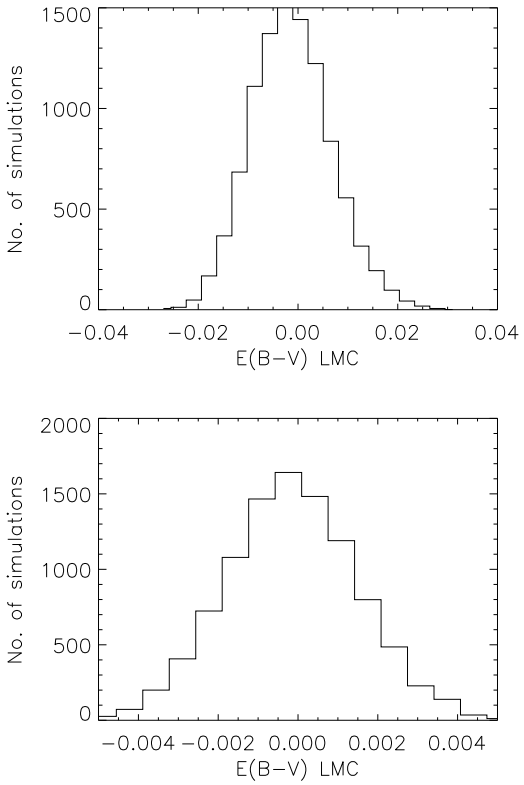

Figure 9. The distribution of apparent values of *E(B−V)* produced by Monte Carlo simulations of 10 000 sets of 37 (top) and 789 (bottom) randomly chosen quasar spectra when shifted randomly to rest frames corresponding to the range in redshift of our absorption systems. See text for additional details. The range of *E(B−V)* covered by the central 68% of simulations provides an estimate of the error on the reddening measured in the absorber composite spectra.

The reddening of each composite was measured by fitting extinction curves of three forms, appropriate for dust in the Milky Way, the Large Magellanic Cloud (LMC) or the Small Magellanic Cloud (SMC) to find the best fitting value of *E(B−V)*. We used the extinction curve tabulations of Cardelli et al. (1989), as extended by O'Donnell (1994), for the Milky Way and of Pei (1992) for the Magellanic Clouds and a total-to-selective extinction ratio $R_V = 3.1$ was assumed throughout.

5.2 Assessment of minimum reddening detectable

Before discussing the results of the reddening analysis, we need to assess the possibility of any systematic effects which may produce a spurious reddening signal. We investigated this with a series of Monte Carlo simulations in which we constructed artificial composites by drawing quasar spectra at random from the reference quasar sample without Ca II and Mg II absorption systems. The selection probability of each spectrum was modified by the redshift path $\Delta z_{\rm quasar}$ over which we searched the spectrum for intervening absorbers:

$$\begin{aligned} \Delta z_{\rm quasar} &= min[1.3, z_{\rm em} - 0.01] \\ &- max[0.84, 1250(1+z_{\rm em})/2175 - 1] \end{aligned} \quad (6)$$

where min and max indicate the minimum and maximum of the values enclosed in the brackets. On selection, each quasar was as-

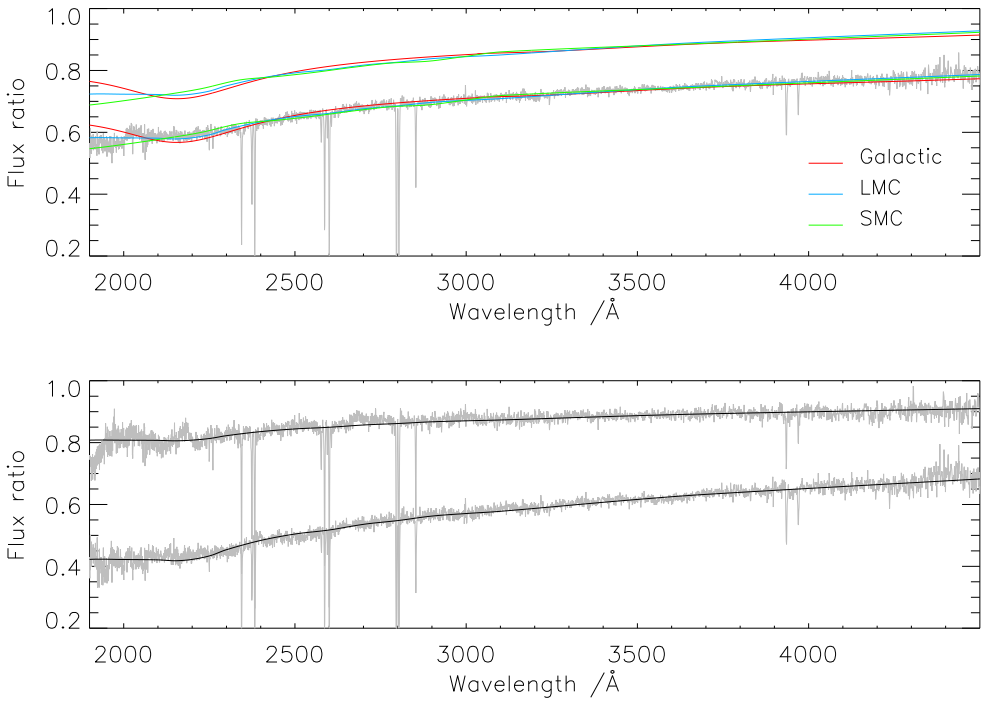

Figure 10. *Top*: Composite spectrum of the 37 quasars with Ca II absorbers after an estimate of the unabsorbed quasar SED has been divided out and each spectrum has been shifted to the rest frame of the absorber before being combined into the composite. The resulting spectrum is fitted with extinction curves appropriate to dust in the Milky Way and the Magellanic Clouds (as indicated) to deduce the values of colour excess, *E(B−V)*, listed in Table 4. The fits to the SED are shown superposed on the spectrum, and are also plotted above it, for clarity. The spectrum and reddening curve normalisation is calculated from the derived *E(B−V)* values and does not play a role during the fitting of the dust reddening curves. *Bottom*: As for the upper panel, but with the Ca II sample now split into the 'High-' and 'Low-$W_{\lambda3935}$' subsamples. The former shows the higher extinction (lower spectrum in the bottom panel). Only the LMC extinction curve fits are shown. See the online version for a colour figure.

signed a random absorption redshift (within the range over which we searched for real absorbers), and shifted to the rest frame of this "absorber". The process was repeated until we had assembled artificial samples of the required size (e.g. 37 for the entire sample of Ca II absorbers).

Composite spectra were then constructed for these artificial absorber samples and analysed for reddening as described above, assuming an LMC extinction curve, and the whole process was repeated 10 000 times. The final product of this exercise are the distributions of apparent *E(B−V)* values which arise simply from variations in the SED of quasars. The two distributions from the simulations designed to mimic the entire Ca II and Mg II-selected DLA samples, are reproduced in Fig. 9. The distributions have a mean *E(B−V)* close to zero and possess a small dispersion, extending to only a few thousandths of a magnitude for the large Mg II sample and to a few hundredths of a magnitude for the Ca II absorbers. In none of the 10 000 simulations performed do we find values of *E(B−V)* greater than 0.031 and 0.0065 for the 37 Ca II absorbers and 789 Mg II-selected DLAs respectively. By taking the 68th percentiles of these distributions, we obtain estimates of the errors on the reddening measured in the absorber samples for use in the analysis of the results.

5.3 Results: reddening by Ca II and Mg II absorbers

The results of the reddening analysis of the real absorbers are collected in Table 4 and illustrated in Figs. 10 and 11. The findings of Paper I in detecting an unequivocal reddening signal from the Ca II absorbers are confirmed by these results. We further strengthen the conclusion that among this sample there is considerable latitude of dust content, with a very pronounced difference between the 'High-' and 'Low-$W_{\lambda3935}$' subsamples (see lower panel of Fig. 10). For both the 'All' and 'High-$W_{\lambda3935}$' samples reddening values of the observed magnitude are never measured in the corresponding Monte Carlo simulations, making these results significant at the >99.99% confidence level. The small *E(B−V)* deduced for the 'Low-$W_{\lambda3935}$' subsample, *E(B−V)*$= 0.026$, is only marginally significant ($< 3\sigma$) given the results of the Monte Carlo simulations.

Turning to the Mg II-selected DLAs, we see from Fig. 11 that they exhibit a substantially lower reddening signal. For the full sample of 789 absorbers (dot-dash line) we measure *E(B−V)*$=$ 0.008 (LMC) which, while very low, is still significant at the >99.99% confidence level according to the results of our Monte Carlo simulations which suggest 1σ errors of 0.0016.

The best-fit LMC extinction curves for the three subsamples split according to the strength of their strong lines are also shown in Fig. 11 with dashed, dotted and continuous lines corresponding, respectively, to $m = 1.5, 2$ and 3 Å (see Section 5.1). The number of absorbers and value of $E(B-V)_{\rm LMC}$ for each subsample are listed in the bottom right-hand corner of the figure. For the composite of 241 absorbers with the strongest Mg II $\lambda2796$ and Fe II $\lambda2600$ absorption lines ($m = 3$ Å; plotted in grey), we also show the best-fit SMC extinction curve (long dash). It is evident from Fig. 11 that, as is the case for the Ca II absorbers, Mg II-selected DLA candidates exhibit a clear trend of increasing dust content with in-

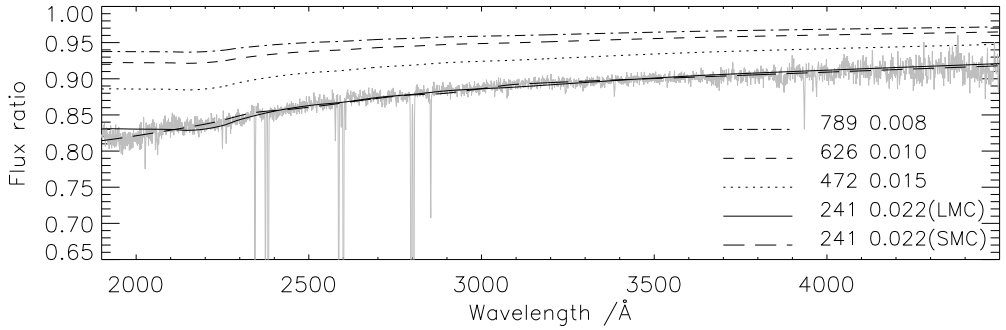

Figure 11. As for Fig. 10 but for Mg II-selected DLA candidates. The curves shown refer to different subsamples of Mg II-selected DLAs ordered by increasing values of $W_{\lambda 2796}$ and $W_{\lambda 2600}$, as described in the text. The full sample includes 789 quasar spectra and shows an average *E(B−V)*$= 0.008$ (for an LMC extinction law), as indicated in the lower right-hand corner of the figure. Corresponding values for the other subsamples are also given. The composite spectrum of the subsample of 241 Mg II-selected DLAs with the highest equivalent widths is shown in grey. Superposed on the spectrum are the best-fitting LMC and SMC extinction curves; both give *E(B−V)*$= 0.022$. The spectrum and reddening curve normalisation is calculated from the derived *E(B−V)* values and does not play a role during the fitting of the dust reddening curves.

Table 4. Estimates of the reddening *E(B−V)* introduced by the Ca II absorbers in the spectra of background quasars. For each reddening curve used, the upper values are for the full 37 absorber sample, and the lower values are for the subsample of 27 absorbers analysed for element column densities. The final two rows give the estimated error for each sample derived from Monte Carlo simulations using an LMC dust curve (see Section 5.2).

	E(B−V)		
Dust law	All	High-$W_{\lambda 3935}$	Low-$W_{\lambda 3935}$
MW	0.057	0.092	0.023
	0.048	0.082	0.012
LMC	0.065	0.103	0.026
	0.055	0.093	0.014
SMC	0.066	0.105	0.026
	0.055	0.095	0.015
err	0.008	0.011	0.011
	0.009	0.013	0.014

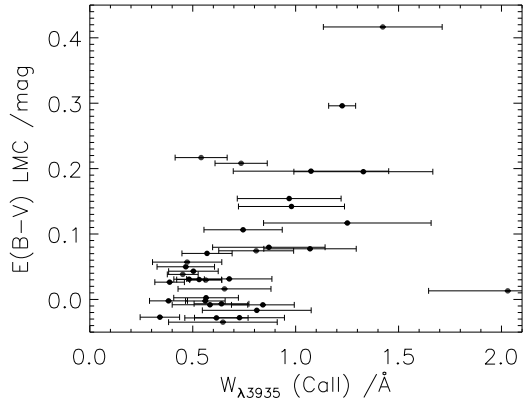

Figure 12. Colour excess due to individual Ca II absorbers plotted as a function of the equivalent width of Ca II $\lambda 3935$.

creasing absorption line equivalent width. Even accounting for the $\sim 57\%$ of sub-DLAs in the sample (Rao 2005), the average dust content of DLAs at these redshifts is tiny. However, for the 30% of absorbers with the highest absorption line equivalent widths, the average *E(B−V)* of the DLAs in the sample may be similar to the average of the Ca II absorbers[5], assuming that sub-DLAs contribute negligible dust reddening.

From Figs. 10 and 11 it can be appreciated that at these mid-UV wavelengths and low levels of extinction there are only subtle differences between the three extinction laws fitted to the data. Even so, a 2175 Å bump as strong as that produced by Galactic dust can probably be excluded for all of the composites. The LMC extinction curve, which flattens near 2250 Å, seems to give the best fit to the reddening produced by the Ca II absorbers. On the other hand, the steeply rising SMC curve appears to be a better fit to the DLA candidates selected via Mg II and Fe II. It will be of interest to establish whether these differences are significant, since they are ultimately related to differences in the size distribution and composition of the dust particles associated with different classes of quasar absorbers.

Finally, while the results of the reddening analysis are most secure when applied to ensembles of absorbers, so that variations due to the intrinsic SEDs of the background quasars are mitigated, it is nevertheless of interest to examine the absorbers' reddening on a case by case basis. The results are also used in Section 6.1 for calculating the obscuration bias inherent in the absorber sample. Each quasar with a Ca II system was divided by the relevant reference spectrum and an LMC extinction curve fitted to the quotient in the rest frame of the absorber to provide individual values of *E(B−V)*. These values are listed in the last column of Tables 1 of the present paper and of Paper I and in Fig. 12 we have plotted them against the equivalent width of Ca II $\lambda 3935$. Despite the scatter, the trend of increasing reddening with equivalent width highlighted by the analysis of the 'High-' and 'Low-$W_{\lambda 3935}$' subsamples can be discerned.

[5] Note that the Ca II absorbers were removed from the Mg II sample and thus do not contribute to these results (Section 2.2).

6 DISCUSSION

We now draw together the results of the previous sections and discuss their implications for the use of Ca II as a tracer of intermediate redshift DLAs, the dust-to-metal ratio of the absorbers and compare their dust contents to high redshift emission-selected galaxies. Firstly, however, we must address the issue of obscuration bias against the dustiest, most metal-rich systems.

6.1 Dust obscuration bias

We begin our discussion with a calculation of the dust obscuration bias against the Ca II systems, necessary for comparison of their true number density to that of DLAs. We further investigate the attenuation which a similar sample of objects would cause at higher redshifts.

There has been much discussion in the literature concerning possible selection bias in magnitude-limited quasar samples against dusty, metal-rich DLAs at high redshift because of the obscuration they would cause to the background quasar light (e.g. Ostriker & Heisler 1984; Fall & Pei 1989, 1993). By selecting quasars at radio wavelengths, which are unaffected by dust, the CORALS survey addressed these concerns (Ellison et al. 2001, 2004). No significant obscuration bias was detected over a wide redshift range. However, the limited statistics of the samples still allow for the possibility that ∼60% of DLAs may remain undetected in optical surveys at intermediate redshifts. In another study, Vladilo & Péroux (2005) used the observed distribution of column densities of Zn II to argue that between 30 and 50% of systems would be obscured, although their conclusions are not borne out by the results of Akerman et al. (2005).

The number of Ca II absorption line systems missing from the quasar sample, as a result of the dimming of the background quasars by dust associated with the Ca II absorber, can be determined by considering the total redshift–path available in the quasar survey with and without intervening absorbers.

A particular quasar contributes a certain redshift–path, Δz (Eq. 6) over which an intervening absorber can be detected. This is determined via the probability of detection, $P(W, z_{abs}, m_i)$, which is a function of equivalent width of the line to be found (W), the absorber redshift and, because fainter objects have lower SNR spectra, the observed i–band magnitude of the quasar (m_i). Summing over all quasars in the sample and over all redshifts for which $P > 0$, produces an estimate of the total absorber redshift–path for the sample for a given equivalent width limit.

Absorbers with an associated reddening, *E(B−V)*, lying in front of a quasar, cause a corresponding observed–frame i–band extinction of A_i mag to the quasar light, making the quasar appear fainter than otherwise would be the case. The redshift–path is now determined via the probability of detection $P(W, z_{abs}, m_i + A_i)$, where m_i is again the unextinguished observed i–band magnitude of the quasar, but the quasar now appears in the sample with a measured i–band magnitude $= m_i + A_i$.

For quasars with magnitudes close to the faint magnitude limit of the sample, the extinction A_i may take the quasar below the faint limit and $P = 0.0$. This reduces the total redshift–path in the sample for which absorbers causing A_i mag of extinction to the background quasar may be identified. Even when the quasar is not lost completely from the sample the probability $P(W, z_{abs}, m_i + A_i) < P(W, z_{abs}, m_i)$, because the SNR of the spectra decreases with increasing magnitude. The extinction due to the intervening absorber thus makes it more difficult to find absorption lines of a given equivalent width.

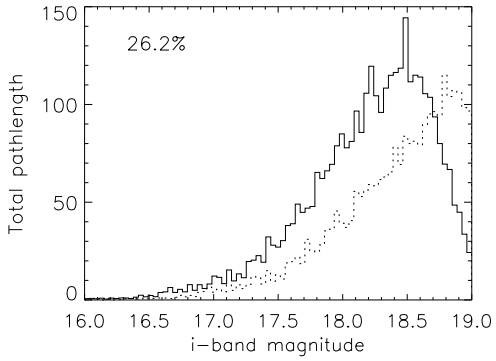

Figure 13. The effect on the total survey redshift–path of dusty absorbers. The continuous histogram shows the observed redshift–path as a function of i-band magnitude of the SDSS quasar sample, while the dotted line shows the effect on the distribution were each quasar line of sight to be intercepted by an absorber with *E(B−V)* = 0.065, $z_{\rm abs} = 0.95$ and $W_{\lambda 3935} = 0.76$ (the mean values of the 37 Ca II absorbers). See the text for further discussion of this figure. The value in the top left is the overall percentage of redshift–path lost.

Fig. 13 shows the redshift–path of the survey available to find an absorber with $z_{\rm abs} = 0.95$, *E(B−V)* = 0.065 (corresponding to $A_i = 0.30$), and $W_{\lambda 3935} = 0.95$ compared to the redshift–path available to find the same absorber if it contained no dust (continuous histogram). In order to calculate the true number of absorbers that would be seen in the absence of any "dust bias", the observed number of such absorbers is multiplied by the factor $\Delta z_{no-dust}/\Delta z_{dust}$.

Calculation of the total number of absorbers missing requires the availability of a well–determined *E(B−V)* distribution. For large values of *E(B−V)* we do not have such a distribution and we choose, conservatively, to calculate the dust obscuration bias only for Ca II absorbers with *E(B−V)*< 0.25. Performing the calculation for the 35 absorbers with *E(B−V)* below this limit, an unbiased number of 58 absorbers is found. Therefore, the obscuration bias towards Ca II absorbers with *E(B−V)*$\lesssim 0.25$ results in a reduction in the observed number of systems seen in the quasar sample of ∼40% ($1 - 35/58$). The $\sim 40\%$ figure is insensitive to binning of the data within sensible ranges of bin size. These 58 absorbers would have an average *E(B−V)*~ 0.1. Because we have simply omitted those absorbers with larger *E(B−V)* from our calculation, these values provide lower limits for the total sample.

As we discuss below, the dust and metal content of the Ca II absorbers is consistent with their identification as DLAs, and their unbiased number density is around 20-30% that of DLAs at similar redshifts. In Section 5.3 we also saw that the majority of intermediate redshift DLAs (that is those selected from their Mg II and Fe II lines) contain negligible quantities of dust. Based on these results, we estimate that dust obscuration in the intermediate redshift DLA population as a whole will cause 8–12% of DLAs with *E(B−V)*$\lesssim 0.25$ to be missed from optical magnitude limited surveys. For this *E(B−V)* limit the results are well within the limits on obscuration bias for DLAs placed by the CORALS survey (Ellison et al. 2004), although an increasing fraction of DLAs with dust contents greater than this limit would be obscured.

Briefly, we note that the reddening effect of intervening

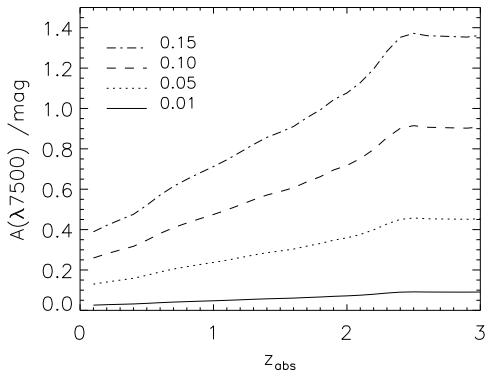

Figure 14. The observed frame (i-band) extinction of quasar light caused by LMC type dust in an intervening galaxy as a function of $z_{\rm abs}$. The different curves are for different columns of dust as parameterised by the *E(B−V)* values given in the top left.

dust on the background quasar SED will affect the selection of quasars for follow-up spectroscopy from the SDSS photometry (Richards et al. 2001). Unfortunately, quantification of this effect is non-trivial at present.

6.1.1 Bias at higher redshifts

To illustrate the impact of extinction caused by intervening dust at higher redshifts on the SDSS quasar sample, Fig. 14 shows the observed frame i-band extinction as a function of $z_{\rm abs}$ caused by an intervening galaxy with reddening $E(B-V)_{\rm LMC}$ = 0.01, 0.05, 0.1 and 0.15. It can be seen that an absorption line system at $z_{\rm abs} = 1$ with $E(B-V)\sim0.1$ and LMC type dust causes just under 0.5 mag of extinction of the quasar light. The same system at $z_{\rm abs} = 2.5$ would cause 0.9 mag of extinction; the effect of dust obscuration is clearly considerably greater at the redshifts probed by traditional DLA studies ($z_{\rm abs} > 2.2$ in the SDSS).

We repeat the dust obscuration analysis for a population of imaginary Ca II absorption systems with the same dust and absorption line properties as our 35 systems with $E(B-V)\lesssim 0.25$, but with redshifts randomly assigned between 2.2 and 3. The quasar sample is taken from the DR3 catalogue of Schneider et al. (2005) with $2.2 < z_{\rm em} < 4$ and $i < 20.2$. Assuming that the identification of DLAs is independent of quasar magnitude (a reasonable approximation for strong absorption features such as a damped Lyman-α line rather than weak Ca II absorption), leads to the prediction that $\sim 75\%$ of the *underlying* population of absorbers with $E(B-V)\lesssim 0.25$ would have been obscured were these observations to have been made at high redshift. The calculation strongly suggests that were a subset of high redshift DLAs to contain small amounts of dust similar to that found in the Ca II absorbers, the vast majority would not be included in current magnitude-limited quasar samples. With the reddening statistics currently available at higher redshifts, a scenario in which 20–30% of DLAs have an average $E(B-V)\gtrsim 0.1$, as found here for intermediate redshifts, will not be easily distinguishable from one in which DLAs are dust free, simply because of the substantial decrease in survey redshift–path for the dusty absorbers.

6.2 The number density of Ca II absorbers and DLAs

Combining the number density per unit redshift of a class of quasar absorption line systems, $n(z)$, with an adopted space density of systems provides a measure of their typical cross-section on the sky. Under the assumption of an unevolving galaxy luminosity function, the frequency with which DLAs occur in quasar spectra implies typical radii of $\sim15\,h^{-1}$ kpc [6]; the corresponding dimensions are $\sim40\,h^{-1}$ kpc for Mg II and Lyman Limit systems, and $\sim70\,h^{-1}$ kpc for C IV absorbers (e.g. Steidel 1993). These inferred sizes are generally in line with the impact parameters observed directly between galaxies and quasar sight-lines in cases where deep imaging has revealed the galaxies responsible for the absorption systems (e.g. Steidel et al. 1994; Adelberger et al. 2005).

Having determined $n(z)$ for Ca II absorbers, we can now place them within this broader context. Specifically, in Section 3 we deduced that the number of Ca II absorbers per unit redshift with a minimum equivalent width of 0.5Å, $n(z, W^{\rm lim} = 0.5)$, is $\sim$0.013. Correcting this value for the incompleteness due to reddening of the background quasars as discussed above, we obtain $n(z, W^{lim} = 0.5) \gtrsim 0.022$ where the lower limit arises from the unconstrained obscuration bias for systems with $E(B-V)\gtrsim 0.25$. Comparing with the values of $n(z)$ for DLAs given by Rao (2005) of 0.079 ± 0.019 at $0.11 < z_{\rm abs} < 0.9$ and 0.120 ± 0.025 at $0.9 < z_{\rm abs} < 1.25$, we conclude that strong Ca II absorption systems have a number density of about 20–30% that of known DLAs.

In the simple model that different classes of absorption line systems arise at different radii in the same galaxies, DLAs occur within the inner regions, while C IV systems arise in outer halos, the sequence being essentially one of decreasing neutral hydrogen column density. The measured value of $n(z)$ for Ca II absorbers then confines them—and their associated reddening—to the innermost 7–8 h^{-1} kpc of galaxies, where the neutral gas column density and/or metallicity are highest. We return to this point in Section 6.4 below.

6.3 Dust-to-metals ratio

We cannot deduce the gas-to-dust ratio applicable to the Ca II systems without a knowledge of their hydrogen column density. However, having measured metal column densities (Section 4) and dust content via the colour excess *E(B−V)* (Section 5), we can determine the dust-to-metals ratio typical of these absorbers. This ratio, which we denote $\mathcal{R}_{\rm DM}$, measures the degree to which refractory elements are incorporated into dust grains which in turn is determined by the balance between dust formation and destruction processes. In a recent study based on the ratio of Zn II to Fe II in a sample of 38 DLAs Vladilo (2004) proposed that $\mathcal{R}_{\rm DM}$ increases with metallicity and, more generally, with the progress of galactic chemical evolution. This is in contrast to most numerical simulations which predict $\mathcal{R}_{\rm DM}$ to be constant with time (e.g. Dwek 1998); measurements from chemically unevolved galaxies at high redshift will provide important constraints for such simulations in the future. At present we simply compare the dust-to-metals ratio of the Ca II absorbers to those found in the local Universe to obtain some clue as to their nature.

As a measure of the dust-to-metals ratio in the present study we define $\mathcal{R}_{\rm DM} \equiv \langle E(B-V)\rangle / \langle N({\rm Zn\ II})\rangle$, adopting Zn (an undepleted Fe-peak element) as an indicator of overall metal content,

[6] Where h is the Hubble constant in units of 100 km s^{-1} Mpc^{-1}

Table 5. Gas-to-dust ratios and metal abundances of the Milky Way, LMC and SMC, along with derived dust-to-metals ratios.

	MW	LMC	SMC
$\langle N(\mathrm{H\ I})\rangle/\langle E(B-V)\rangle$ (10^{22} cm^{-2} mag^{-1})	0.493[a]	2.0[b]	7.9[c]
(Fe/H) ($\times 10^{-5}$)	2.95 [d]	1.12 [e]	0.55 [f]
$\langle E(B-V)\rangle/\langle N(\mathrm{Zn\ II})\rangle$ (10^{-15} mag cm^2)	4.7	3.1	1.6

[a] Diplas & Savage (1994)
[b] Koornneef (1982)
[c] Fitzpatrick & Massa (1990)
[d] Lodders (2003)
[e] B-type stars, Korn et al. (2000)
[f] A-type stars, Venn (1999)

and colour excess as a substitute for dust column density. It is necessary to confine the analysis to the subset of 27 Ca II absorbers at redshifts which include the Zn II lines; their values of N(Zn II) are given in Table 3 and the lower set of values in Table 4 gives *E(B−V)* for this subset. We make no allowance for the incompleteness effects due to dust extinction discussed above, since it is not possible to correct the Zn II column densities without knowing the distribution of individual values. For the three Ca II composites, 'All', 'High-$W_{\lambda3935}$' and 'Low-$W_{\lambda3935}$', we find $\mathcal{R}_{\rm DM} = 8.3^{+1.9}_{-1.8}$, 6.4 ± 1.2 and $3.6^{+3.7}_{-3.6} \times 10^{-15}$ mag cm^2 respectively. The errors have been calculated by propagation of those quoted in Tables 3 and 4 and the LMC *E(B−V)* values have been used. Although at first glance there appears to be a discrepancy between the values of the three samples, with the mean value being larger than both the two subsamples, the errors are such that this is not significant. Future SDSS releases, or follow up observations of Zn II in individual absorbers, may improve the situation. However, the current measurement errors clearly do not allow us to distinguish between the 'High-' and 'Low-$W_{\lambda3935}$' samples.

We have also determined corresponding values of $\mathcal{R}_{\rm DM}$ appropriate to the interstellar media of the Milky Way, LMC and SMC from their metallicities (Fe/H) and average gas-to-dust ratios $\langle N(\mathrm{H\ I})\rangle/\langle E(B-V)\rangle$, assuming a solar (Zn/Fe) ratio ($= 1.45 \times 10^{-3}$, Lodders 2003) throughout:

$$\frac{\langle E(B-V)\rangle}{\langle N(\mathrm{Zn\ II})\rangle} = \left[\frac{\langle N(\mathrm{H\ I})\rangle}{\langle E(B-V)\rangle} \times \frac{\mathrm{Fe}}{\mathrm{H}} \times \left(\frac{\mathrm{Zn}}{\mathrm{Fe}}\right)_{\odot}\right]^{-1} \quad (7)$$

We have chosen this approach because the quantities on the right-hand side of Eq. 7 are much better known than the ratios *E(B−V)*/N(Zn II) for individual sight-lines, especially to stars in the Magellanic Clouds. The results are collected in Table 5, where we also give references to the sources of the measurements. For the Magellanic Clouds, we have adopted the Fe abundance determined in young B- and A-type stars—other abundance indicators, including Zn, are consistent with these values (Hill 2004).

The sequence of decreasing dust-to-metals ratios from the Milky Way to the LMC and SMC evident in Table 5 is in qualitative agreement with previous measurements (Issa et al. 1990). It also appears to be in accord with Vladilo's (2004) proposal of a metallicity dependence of $\mathcal{R}_{\rm DM}$, although we should bear in mind that the Fe abundance in the Magellanic Clouds has so far been determined in only a few objects and may be subject to revision in future. Different star formation rates may also effect $\mathcal{R}_{\rm DM}$ (Lisenfeld & Ferrara 1998).

It is of considerable note that we find that the values determined above for the Ca II absorbers, $\mathcal{R}_{\rm DM} = (4-8) \times 10^{-15}$ mag cm^2, are as high, or higher, than those derived for the Milky Way, particularly for the 'High-$W_{\lambda3935}$' sample. Previous estimates for the dust-to-metals ratios in high redshift DLAs have been obtained by comparison of the level of depletion of elements to that seen in the Milky Way ISM (Pettini et al. 1994, 1997), suggesting average ratios of $\sim$1/2 those of the Milky Way. We saw earlier (Fig. 7) that the ratios of Cr, Fe and Ti to Zn[7] are typically one third of solar in the Ca II absorbers, suggesting that two thirds of these refractory elements have been incorporated into the dust. Given that in the Milky Way the dust depletions of these elements are $\gtrsim$90%, we would have expected values of $\mathcal{R}_{\rm DM}$ of about 2/3 that of the Milky Way, i.e.$\approx 3 \times 10^{-15}$ mag cm^2. Given the uncertainties in our derivation of this ratio, the disagreement between element depletions and dust-to-metals ratios may not be too serious. Only further observations will clarify this issue.

6.4 The H I column densities of Ca II absorbers

Ultimately, we need the neutral hydrogen column densities of the Ca II absorbers in order to fully understand their relation to other classes of quasar absorption systems. Unfortunately, this will not be possible for at least several years, given the current lack of spaceborne UV spectrographs. We therefore have to rely on indirect arguments based on known relations between the column density of elements, dust content and N(H I) in the local universe and in confirmed DLAs.

Lower limits to the values of N(H I) appropriate to our samples of Ca II absorbers can be derived from their Zn II column densities assuming a solar abundance of Zn. From the values in Table 3 we obtain $\log[N(\mathrm{H\ I})] \geqslant 20.19$, 20.53 and 19.96 for the entire sample, and 'High-' and 'Low-$W_{\lambda3935}$' subsamples respectively. Thus, even without knowledge of their metallicity, it can be seen immediately that the strongest ($W_{\lambda3935} > 0.68$) and, as we have shown in Section 5.3, dustiest Ca II systems are very likely to be DLAs. Even the lower equivalent width sample is on average only within a factor of two of the canonical DLA limit $\log[N(\mathrm{H\ I})] = 20.3$.

The assumption of solar metallicity is likely to be overly conservative in the present context. The abundance of Zn has been measured in many DLAs at $z_{\rm abs} \simeq 1$ and found to be generally sub-solar. The recent compilation by Kulkarni et al. (2005) shows a mean and median [Zn/H]$\simeq -1.0$ at these intermediate redshifts. If the Ca II absorbers were randomly chosen from known DLA samples, then their corresponding values of $\log[N(\mathrm{H\ I})]$ would be ten times higher than those given above, placing them firmly at the upper end of the distribution of column densities of DLAs.

A second constraint on N(H I) can be obtained from consideration of the gas-to-dust ratios, $\langle\ N(\mathrm{H\ I})\ \rangle/\langle\ E(B-V)\rangle$, in the Milky Way and the Magellanic Clouds given in Table 5. Again, the Milky Way value will give a lower limit to N(H I) for a given reddening *E(B−V)*, if 1) the gas-to-dust ratio increases with decreasing metallicity, as suggested by local trends (Table 5) and 2) the Ca II absorbers have lower metallicities than the Milky Way, as would generally be assumed for $z \sim 1$ galaxies. The values of *E(B−V)* in the

[7] Mn is also found to be underabundant, but this could be due to combined dust depletion and intrinsic abundance effects (Ledoux et al. 2002).

second row of Table 4 and the Milky Way gas-to-dust ratio imply $\log[N(\mathrm{H\ I})] \geqslant 20.38$, 20.61, and 19.78 for the three samples with Zn II measurements. The corresponding values for the full sample of 37 Ca II systems (top row of Table 4) would be somewhat higher; correcting for the incompleteness caused by dust obscuration up to $E(B-V) \lesssim 0.25$, as discussed in Section 6.1, would lead to $\log[N(\mathrm{H\ I})] \geqslant 20.69$ for Ca II absorbers in general. The conclusion of this discussion is that, given their column densities of metals and dust, Ca II absorption systems with $W_{\lambda 3935} > 0.68$ Å are highly likely to be DLAs, and it is certainly possible that all absorbers presented here are DLAs or close to the nominal DLA column density limit.

6.4.1 The highest column density DLAs?

We saw earlier (Section 6.2) that the number density $n(z)$ of Ca II absorbers with $W_{\lambda 3935} > 0.5$ Å is 20–30% that of DLAs after correcting for dust obscuration bias. Possible interpretations of this statistic, together with the conclusions above, are that Ca II systems trace the subset of DLAs with: (i) the largest values of neutral hydrogen column density; (ii) the highest metallicities; or (iii) the largest volume densities n_{H}—or a combination of all three effects. The third possibility results from the fact that, being a minor ionisation stage of Ca, the fraction of Ca which is singly ionised grows in proportion to n_{H}^2. Thus, two clouds with the same total column density N(H I) and the same Ca abundance, but differing volume densities n_{H}, will have different column densities of Ca II, due to the higher recombination coefficient for Ca II in the denser cloud (Hobbs 1974).

It is difficult to assess the relative importance of these three possibilities without additional information, although the third one must surely contribute, given the tendency of most Ca II systems to have stronger Mg I $\lambda 2853$ absorption for a given Mg II $\lambda 2796$ equivalent width (see Fig. 2)—neutral Mg is also a trace state in the ISM of galaxies.

We have explored the consequences of option (i) by taking as a starting point the column density distribution $f[N(\mathrm{H\ I})]$ of DLAs in the SDSS DR3 catalogue from Prochaska et al. (2005). With Prochaska et al.'s double power law parameterisation for DLAs in the redshift interval $2.5 < z_{\mathrm{abs}} < 3$ the 20% of DLAs with the largest values of N(H I) have $20.95 < \log[N(\mathrm{H\ I})] < 22$, and $\langle \log[N(\mathrm{H\ I})] \rangle = 21.25$. This value is insensitive to the exact upper limit assumed, and also varies little with the precise redshift dependent parameterisation used. The mean value would imply a metallicity [Zn/H] $= -1.06$ and a gas-to-dust ratio $\langle N(\mathrm{H\ I}) \rangle / \langle E(B-V) \rangle = 3.2 \times 10^{22}$ cm^{-2} mag^{-1} for the subsample of observed Ca II absorbers with spectra covering the Zn II $\lambda 2026$ region and assuming LMC type dust for the latter value. In this case, the metallicity would be typical of DLAs in general and in line with the depletion trends in [Cr/Zn] (Fig. 7). The gas-to-dust ratio would be intermediate between those of the LMC and SMC (see Table 5).

These parameters seem plausible, however, they are at odds with the idea that the dust-to-metals ratio should be lower at low metallicities, as proposed by Vladilo (2004), given that in Section 6.3 we found $\mathcal{R}_{\mathrm{DM}}$ to be as high, or higher, than in the Milky Way ISM. If the proposal of Vladilo (2004) were to be correct, our detection of reddening by the Ca II absorbers argues in favour of the hypothesis that they trace DLAs with an unusually high dust content, due to either high metallicities or densities (or both), rather than simply DLAs at the upper end of the distribution of N(H I). The results of Ledoux et al. (2003), that the molecular hydrogen column density of DLAs is not correlated with N(H I), but does imply high metallicities and depletions, is also consistent with the scenario that the dusty Ca II absorbers have higher metallicities rather than higher column densities.

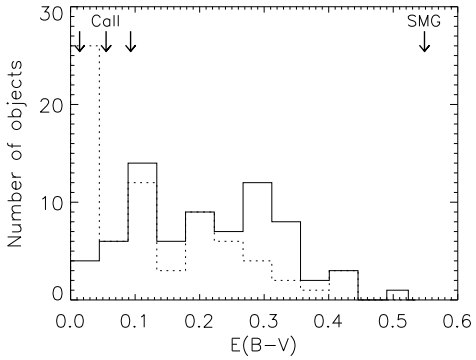

Figure 15. The distribution of *E(B−V)* for UV-selected galaxies at $z \sim 2$ from Shapley et al. (2005). The solid and dotted histograms show the best-fitting *E(B−V)* for a constant or bursting star forming model respectively. Arrows indicate the average *E(B−V)* values of the observed Ca II absorbers, uncorrected for incompleteness effects, and the average value for the sub-millimetre selected galaxies of Borys et al. (2005).

6.5 Dust in high redshift galaxies

We conclude our discussion by comparing the reddening found here for the Ca II absorption-selected galaxies with measurements for high redshift galaxies detected directly via their stellar or dust emission. In two independent studies, Shapley et al. (2001) and Papovich et al. (2001) concluded that Lyman break galaxies at $z \sim 3$ have *E(B−V)*$= 0-0.4$, and a median $\langle E(B-V) \rangle \simeq 0.15$ by comparing their observed rest frame UV-to-optical colours to model spectral energy distributions. The addition of rest frame near-IR photometry made possible by the *Spitzer* satellite has led to similar estimates for star-forming galaxies at $z \sim 2$ (Shapley et al. 2005) shown in Fig. 15. Borys et al. (2005) performed an equivalent analysis on a sample of sub–millimetre selected galaxies, finding these particular SCUBA sources to have stellar masses similar to the most massive of the UV–selected galaxies and an average extinction in the V–band, A_V, of 1.7 equivalent to an *E(B−V)* of 0.55 assuming a LMC dust curve. This value is also indicated in the figure.

While it is not surprising to find that the highest values of *E(B−V)* apply to galaxies that are luminous at far-IR/sub-mm wavelengths, where we see directly the emission by dust, it is noteworthy that the more normal star-forming galaxies selected via their rest frame UV stellar light are also generally more reddened than the Ca II absorbers. An obvious question regarding the differences in the reddening between absorption- and UV-selected galaxies is the effect of dust obscuration bias on the former sample. It is clearly the case, given the calculations of dust bias presented in this chapter, that the majority of absorbers with dust contents as high as the upper end of the UV–selected galaxy distribution will be obscured from sight. Whether such absorbers represent a significant population depends primarily on the total cross section on the sky with column densities of dust this large. A second complication in making a direct comparison between the effect of extinction on the absorption- and UV-selected systems is that the reddening curves applicable to the derivation of *E(B−V)* are very different for the

two cases, because of the geometrical configuration of dust and star formation in galaxies.

Nevertheless, until now, the two populations of absorption- and emission-selected galaxies have been differentiated by dust content, as well as metallicity and star formation rates. It is thus of potential importance that the mean $\langle E(B-V)\rangle \gtrsim 0.1$ which we deduce for the Ca II absorbers after correcting for the missing systems in our magnitude limited quasar survey is not substantially lower than the median $E(B-V) = 0.15 - 0.20$ of star-forming galaxies at $z = 2 - 3$. Possibly, Ca II absorbers are an intermediate link in a sequence of star formation rates, metallicities and dust content which stretches from the relatively quiescent, metal- and dust-poor DLAs to the actively star-forming, metal and dust-rich Lyman break galaxies. Only further observations will provide the information needed to confirm or refute this.

7 SUMMARY AND CONCLUDING REMARKS

In this paper we have extended the study of intermediate redshift ($z_{\rm abs} \sim 1$) absorption systems selected via Ca II equivalent width which was begun by Wild & Hewett (2005b) by: (a) expanding the sample of Ca II absorbers to a total of 37 with new systems selected from the SDSS DR4; (b) measuring the relative abundances of refractory elements from composite spectra of a subset of 27 systems at redshifts such that the Zn II $\lambda\lambda 2026, 2062$ doublet is covered in the SDSS spectra; and (c) comparing the degree of reddening imposed on the spectra of background quasars by dust in intervening Ca II absorbers to that caused by Mg II-selected DLAs.

We have confirmed the detection of of significant reddening, at the level of $E(B-V) \simeq 0.06$, associated with the Ca II absorption systems. This is particularly noteworthy considering that much larger samples of Mg II-selected DLAs at the same redshifts and of Lyman α-selected DLAs at higher redshifts have all failed so far to show a clear reddening signal. Only Mg II-selected DLAs with the strongest metal lines approach the reddening of the Ca II absorbers, once allowance is made for the fraction of sub-DLAs in the sample.

Given the lack of direct measurements of N(H I) for the foreseeable future, we have considered various lines of indirect evidence which all point towards the conclusion that our absorption systems with a Ca II $\lambda 3935$ equivalent width $W_{\lambda 3935} \gtrsim 0.68$ Å are a subset of the DLA population, and those with lower equivalent widths are likely to be DLAs or lie close to the nominal DLA column density limit:

(i) The strengths of Mg II $\lambda 2796$ and Fe II $\lambda 2600$ absorption lines fulfil the DLA selection criteria of Rao (2005) in all but two cases (which are borderline—see Fig. 2).

(ii) The ratios of the refractory elements Cr, Fe, Ti and Mn to the undepleted Zn all span similar ranges to those in confirmed DLAs, with the strongest Ca II systems exhibiting some of the most pronounced degrees of depletion encountered so far (Fig. 7).

(iii) With the conservative assumption of solar metallicity, the measured Zn II column densities imply an average $\log[N(\text{H I})] = 20.19$, just below the threshold of $\log[N(\text{H I})] = 20.3$ considered to be the definition of a damped system. Even a moderate, and more realistic, underabundance of Zn and other Fe-peak elements in these systems would bring them within the conventional DLA samples.

(iv) In itself, the magnitude of reddening detected implies $\log[N(\text{H I})] \geqslant 20.43$ adopting the conservative assumption of a Galactic gas-to-dust ratio.

Furthermore, these are likely to be underestimates for the population of Ca II absorbers as a whole, since our magnitude limited quasar survey certainly misses a greater fraction of absorbers with the largest values of $E(B-V)$.

After correcting for our detection efficiency, including dust obscuration bias, we calculate that Ca II absorbers with $W_{\lambda 3935} > 0.5$ Å have a number density per unit redshift $n(z) \sim 0.022$ at $\langle z_{\rm abs}\rangle = 0.95$, 20–30% of that of Lyman α-selected DLAs at the same redshifts. This in turn implies that the regions within galaxies giving rise to strong Ca II absorption have small dimensions, with radii of 7–8 h^{-1} kpc. Possibly, these are the inner regions where the highest column densities N(H I), gas densities $n_{\rm H}$, and/or metallicities (Zn/H) are encountered and where the characteristics of DLAs begin to approach those of the galaxies detected directly via their UV stellar light.

Even without a knowledge of their neutral hydrogen column densities we can deduce the typical dust-to-metals ratio $\mathcal{R}_{\rm DM}$ of Ca II-selected DLAs from their measured columns of metals (via Zn II) and dust [via $E(B-V)$]. We find high values of $\mathcal{R}_{\rm DM}$, as high or higher than the typical ratio in the Milky Way, although the errors affecting our estimates are still large. This may be a further indication that Ca II absorbers are at an advanced stage of chemical evolution if the dust-to-metals ratio correlates with metallicity.

Interestingly, we find a clear trend of increasing reddening and element depletions with Ca II equivalent width; a positive correlation of $E(B-V)$ with line strength is also present among the Mg II-selected DLA candidates. Since these strong lines are generally saturated, their equivalent widths depend more on the velocity dispersion of the absorbers than on their column densities. Thus, these trends may indicate that the systems with the larger values of $E(B-V)$ are either more massive or undergoing some form of disturbance.

Whatever their exact nature, it is clear that Ca II-selected systems are an interesting class of quasar absorbers which have been relatively neglected until now and yet merit considerable attention. Their relation to the DLA population at large will be clarified by measuring the Ca II lines associated with confirmed DLAs; recent improvements in detector and spectrograph efficiency at far-red and near-IR wavelengths now make the Ca II $\lambda\lambda 3935, 3970$ lines more easily accessible to high resolution spectroscopy at redshifts $z_{\rm abs} > 1$. Such measurements will establish the relation between N(H I), Ca II equivalent width and dust content, and will provide firm estimates of the metallicities and gas-to-dust ratios which are lacking from the current analysis.

At $z \sim 1$ imaging is considerably easier than for the bulk of known DLAs at $z > 2$ and it should be relatively straightforward to identify the objects responsible for the Ca II absorption. In this respect, our sample of 37 Ca II absorbers represents a major injection of new data into a field of study which up to now has been severely limited by the small number of known DLAs at $z \sim 1$. The images will establish whether the absorption does take place in the inner regions of galaxies which have undergone significant chemical evolution, as we propose, or whether DLAs with Ca II absorption single out major mergers (Bowen 1991). Future papers in this series will investigate further both the relation between Ca II absorbers and DLAs, and the properties of the host galaxies of Ca II absorbers through follow-up observations.

Ultimately, the combination of absorption line spectroscopy and deep imaging will clarify the link between absorption- and emission-selected galaxies and thereby help constrain galaxy evolution models. The Ca II absorbers may well turn out to be the key

to understanding how different classes of high redshift galaxies fit into a unified picture.

ACKNOWLEDGMENTS

We would like to thank Chris Akerman, Dave Alexander, Sara Ellison and Michael Murphy for valuable discussions, preprints of papers and electronic data tables. VW acknowledges the award of a PPARC research studentship. This work made extensive use of the Craig Markwardt IDL library.

Funding for the Sloan Digital Sky Survey (SDSS) has been provided by the Alfred P. Sloan Foundation, the Participating Institutions, the National Aeronautics and Space Administration, the National Science Foundation, the U.S. Department of Energy, the Japanese Monbukagakusho, and the Max Planck Society. The SDSS Web site is http://www.sdss.org/. The SDSS is managed by the Astrophysical Research Consortium (ARC) for the Participating Institutions. The Participating Institutions are The University of Chicago, Fermilab, the Institute for Advanced Study, the Japan Participation Group, The Johns Hopkins University, Los Alamos National Laboratory, the Max-Planck-Institute for Astronomy (MPIA), the Max-Planck-Institute for Astrophysics (MPA), New Mexico State University, University of Pittsburgh, Princeton University, the United States Naval Observatory, and the University of Washington.

References

Adams W. S., 1949, ApJ, 109, 354
Adelberger K. L., Shapley A. E., Steidel C. C., Pettini M., Erb D. K., Reddy N. A., 2005, ApJ, 629, 636
Adelman-McCarthy J. K., et al. (The SDSS Collaboration), 2005, astro-ph/050711
Akerman C. J., Ellison S. L., Pettini M., Steidel C. C., 2005, A&A, in press (astro-ph/0506180)
Blades J. C., 1988, in , Blades J. C., Turnshek D. A., & Norman C., eds, QSO absorption lines: Probing the universe, Cambridge University Press, p. 147
Borys C., Smail I., Chapman S. C., Blain A. W., Alexander D. M., Ivison R. J., 2005, ApJ, in press (astro-ph/0507610)
Bowen D. V., 1991, MNRAS, 251, 649
Bowen D. V., Pettini M., Penston M. V. & Blades C., 1991, MNRAS, 251, 649
Cardelli J. A., Clayton G. C., Mathis J. S., 1989, ApJ, 345, 245
Carilli C. L., Menten K. M., Reid M. J., Rupen M. P., Yun M. S., 1998, ApJ, 494, 175
Connolly A. J., Szalay A. S., 1999, AJ, 117, 2052
Connolly A. J., Szalay A. S., Bershady M. A., Kinney A. L., Calzetti D., 1995, AJ, 110, 1071
Diplas A., Savage B. D., 1994, ApJ, 427, 274
Dwek E., 1998, ApJ, 501, 643
Ellison S. L., Churchill C. W., Rix S. A., Pettini M., 2004, ApJ, 615, 118
Ellison S. L., Hall P. B., Lira P., 2005, AJ, in press (astro-ph/0507418)
Ellison S. L., Yan L., Hook I. M., Pettini M., Wall J. V., Shaver P., 2001, A&A, 379, 393
Fall S. M., Pei Y. C., 1989, ApJ, 337, 7
Fall S. M., Pei Y. C., 1993, ApJ, 402, 479
Fitzpatrick E. L., Massa D., 1990, ApJS, 72, 163
Francis P. J., Hewett P. C., Foltz C. B., Chaffee F. H., 1992, ApJ, 398, 476
Glazebrook K., Offer A. R., Deeley K., 1998, ApJ, 492, 98
Hall P. B., Richards G. T., York D. G., Keeton C. R., Bowen D. V., Schneider D. P., Schlegel D. J. & Brinkmann J., 2002, ApJL, 575, L51
Hewett P. C., Foltz C. B., 2003, AJ, 125, 1784
Hewett P. C., Irwin M. J., Bunclark P., Bridgeland M. T., Kibblewhite E. J., He X. T., Smith M. G., 1985, MNRAS, 213, 971
Hill V., 2004, in McWilliam A., Rauch M., eds, Origin and Evolution of the Elements p. 205
Hobbs L. M., 1974, ApJL, 188, L107
Issa M. R., MacLaren I., Wolfendale A. W., 1990, A&A, 236, 237
Kendall M. G., 1975, Multivariate Analysis. Griffen, London
Koornneef J., 1982, A&A, 107, 247
Korn A. J., Becker S. R., Gummersbach C. A., Wolf B., 2000, A&A, 353, 655
Kulkarni V. P., Fall S. M., Lauroesch J. T., York D. G., Welty D. E., Khare P., Truran J. W., 2005, ApJ, 618, 68
Ledoux C., Bergeron J., Petitjean P., 2002, A&A, 385, 802
Ledoux C., Petitjean P., Srianand R., 2003, MNRAS, 346, 209
Lisenfeld U., Ferrara A., 1998, ApJ, 496, 145
Lodders K., 2003, ApJ, 591, 1220
Madgwick D. S., Lahav O., Baldry I. K., et al. (The 2dFGRS Team), 2002, MNRAS, 333, 133
Marschall L. A., Hobbs L. M., 1972, ApJ, 173, 43
Morton D. C., 2003, ApJS, 149, 205
Murphy M. T., Liske J., 2004, MNRAS, 354, L31
Murtagh F., Heck A., 1987, Multivariate data analysis. Astrophysics and Space Science Library, Dordrecht: Reidel, 1987
Nestor D. B., Rao S. M., Turnshek D. A., Vanden Berk D., 2003, ApJL, 595, L5
Nestor D. B., Turnshek D. A., Rao S. M., 2005, ApJ, 628, 637
O'Donnell J. E., 1994, ApJ, 422, 158
Ostriker J. P., Heisler J., 1984, ApJ, 278, 1
Papovich C., Dickinson M., Ferguson H. C., 2001, ApJ, 559, 620
Pei Y. C., 1992, ApJ, 395, 130
Petitjean P., Aracil B., Srianand R. & Ibata R., 2000, A&A, 359, 457
Pettini M., 2004, in Esteban C., Garci Lopez R. J., Herrero A., Sanchez F., eds, Cosmochemistry. The melting pot of the elements Cambridge University Press, p. 257
Pettini M., Boksenberg A., Hunstead R. W., 1990, ApJ, 348, 48
Pettini M., Ellison S. L., Steidel C. C., Bowen D. V., 1999, ApJ, 510, 576
Pettini M., Ellison S. L., Steidel C. C., Shapley A. E., Bowen D. V., 2000, ApJ, 532, 65
Pettini M., King D. L., Smith L. J., Hunstead R. W., 1997, ApJ, 478, 536
Pettini M., Smith L. J., Hunstead R. W., King D. L., 1994, ApJ, 426, 79
Prochaska J. X., Herbert-Fort S., Wolfe A. M., 2005, ApJ, in press (astro-ph/0508361)
Prochaska J. X., Wolfe A. M., Tytler D., Burles S., Cooke J., Gawiser E., Kirkman D., O'Meara J. M., Storrie-Lombardi L., 2001, ApJS, 137, 21
Rao S. M., 2005, IAU Colloquium 199 (astro-ph/0505479)
Rao S. M., Turnshek D. A., 2000, ApJS, 130, 1
Reichard T. A., Richards G. T., Hall P. B., Schneider D. P., Vanden Berk D. E., Fan X., York D. G., Knapp G. R., Brinkmann J., 2003, AJ, 126, 2594

Reichard T. A., Richards G. T., Schneider D. P., Hall P. B., Tolea A., Krolik J. H., Tsvetanov Z., Vanden Berk D. E., York D. G., Knapp G. R., Gunn J. E., Brinkmann J., 2003, AJ, 125, 1711
Richards G. T., Fan X., Schneider D. P., et al., 2001, AJ, 121, 2308
Roweis S., 1997, Neural Information Processing Systems, 10, 626
Savage B. D., Sembach K. R., 1996, ARAA, 34, 279
Schneider D. P., Hall P. B., Richards G. T., et al. 2005, astro-ph/0503679
Shapley A. E., Steidel C. C., Adelberger K. L., Dickinson M., Giavalisco M., Pettini M., 2001, ApJ, 562, 95
Shapley A. E., Steidel C. C., Erb D. K., Reddy N. A., Adelberger K. L., Pettini M., Barmby P., Huang J., 2005, ApJ, 626, 698
Steidel C. C., 1993, in Shull J. M., Thronson H. A., eds, ASSL Vol. 188: The Environment and Evolution of Galaxies Dordrecht: Kluwer, p. 263
Steidel C. C., Pettini M., Dickinson M., Persson S. E., 1994, AJ, 108, 2046
Venn K. A., 1999, ApJ, 518, 405
Vladilo G., 2004, A&A, 421, 479
Vladilo G., Péroux C., 2005, astro-ph/0502137
Welty D. E., Hobbs L. M., Lauroesch J. T., Morton D. C., Spitzer L., York D. G., 1999, ApJS, 124, 465
Wakker B. P., Mathis J. S., 2000, ApJS, 544, L107
Wang T. G., Dong X. B., Zhou H. Y. & Wang J. X., 2005, ApJL, 622, L101
Weymann R. J., Morris S. L., Foltz C. B., Hewett P. C., 1991, ApJ, 373, 23
Wild V., Hewett P. C., 2005a, MNRAS, 358, 1083
Wild V., Hewett P. C., 2005b, MNRAS, 361, L30
Wolfe A. M., Prochaska J. X., Gawiser E., 2005, ARAA, in press
Wolfe A. M., Turnshek D. A., Smith H. E., Cohen R. D., 1986, ApJS, 61, 249
Yip C. W., Connolly A. J., Szalay A. S., et al. 2004, AJ, 128, 585
Zwaan M. A., van der Hulst J. M., Briggs F. H., Verheijen M. A. W., Ryan-Weber E., 2005, MNRAS, in press

APPENDIX A: IDENTIFYING BROAD ABSORPTION LINE QUASARS

BAL quasars are known to make up around 15% of all active galactic nuclei at $z_{\rm em} > 1$ (Hewett & Foltz 2003; Reichard et al. 2003). They are characterised by broad absorption troughs blueward of emission lines of many ionisation states—presumably arising in outflowing material—and an overall reddening of the quasar spectral energy distribution. While much effort is being devoted to understanding the origin of these spectral features and the position of BAL quasars in the unified model of active galactic nuclei, their generally redder SEDs are a potential source of additional uncertainty in studies of reddening caused by intervening absorbers. Similarly associated absorption systems, in which hydrogen and metal absorption lines seen in the spectra of quasars are associated with the quasar host galaxy itself, also appear in quasars with generally redder SEDs (Carilli et al. 1998). Thus, we need to remove quasars whose spectra show evidence of BALs, or strong associated absorption, from the sample prior to our dust reddening analysis. Due to the large number of quasar spectra in the SDSS, the problem becomes one of developing suitable automatic pattern recognition techniques capable of coping with the wide range in strength and shape of the absorption features.

The main difficulty lies in defining the quasar continuum around the emission lines. Traditionally, this has been achieved by fitting power laws to the spectra away from regions containing emission lines. Reichard et al. (2003) compared this method to the use of composite spectra in the SDSS early data release; they also provide a history of BAL detection and an extensive discussion of the different methods which will not be repeated here. Here, we present a third method which makes use of a principal component analysis (PCA) of the spectra to reconstruct the quasar continua and emission lines, without the broad absorption features. The main advantage of PCA over the use of composite spectra for reconstructing the quasar continua is the tailoring of the fit to each individual object; all the detailed information which is lost through power law fits can also be retained.

A1 Principal Component Analysis

PCA has become a standard spectral analysis technique in astronomy, generally used to automatically classify stars, galaxies and quasars (e.g. Francis et al. 1992; Connolly et al. 1995; Madgwick et al. 2002; Yip et al. 2004), but also in data reduction (Glazebrook et al. 1998; Wild & Hewett 2005a) and reconstruction problems. Mathematical details can be found in standard texts on multivariate analysis (e.g. Kendall 1975; Murtagh & Heck 1987), but the main aspects of the technique can be easily explained. An array of N spectra each containing M pixels can alternatively be visualised as each spectrum being a point in an M dimensional space. PCA searches for the lines of greatest variance in the cloud of points representing the spectra, each line being orthogonal to all the previous ones, therefore picking out a basis set in which the eigenvectors (principal components or 'eigenspectra') are ordered according to their relative contribution to the overall variance in the array. PCA has no knowledge of physical information; consequently physical features which occur in only a few spectra will be lost in the noise and 'features' resulting from low SNRs may be erroneously identified as signal. The former property is advantageous when identifying the small fraction of spectra exhibiting strong absorption.

Once the eigenspectra are created, any suitable quasar spectrum can be reconstructed by projection onto the eigenspectra. The reconstructed spectrum will only contain the components present in the eigenspectra and a poor fit will be obtained for quasar spectra with shape and/or features not well represented in the input sample.

A2 Details of the method

Given the size of the SDSS catalogue, it is possible to create high SNR eigenspectra for different subsamples of quasars. Splitting the sample into redshift bins can greatly simplify the PCA analysis by minimising the 'missing data' caused by differing rest wavelength coverage. We defined three redshift bins such that all the spectra contained within them have less than 25% of their pixels missing when placed onto a common rest frame wavelength array: $0.823 < z_{\rm em} < 1.537$, $1.405 < z_{\rm em} < 2.179$ and $2.172 < z_{\rm em} < 3.193$. We chose to have a slight overlap between the redshift bins in order to increase the number of spectra contributing to either end of the eigenspectra, and thus improve the SNR at the ends of the eigenspectra; a small fraction of the spectra therefore contribute to two bins. We also limited our analysis to wavelengths longwards of 1250 Å, i.e. away from the Lyman-α forest and the peak of the Lyman-α emission in the quasars.

Several steps are involved before the final sample of BAL quasars was produced. The first step was to identify strong BAL

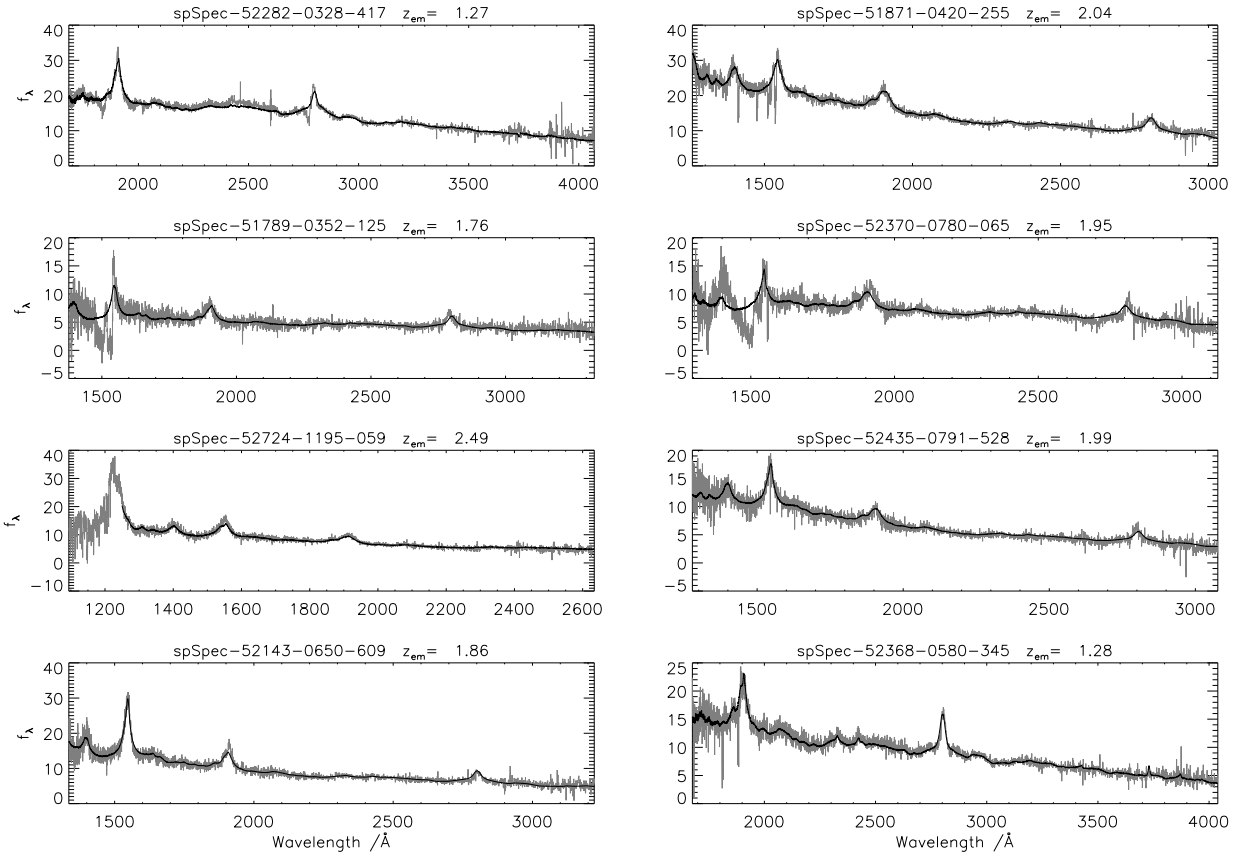

Figure A1. Examples of quasars with (upper four panels) and without (lower four panels) broad absorption troughs and/or intrinsic absorption lines; the PCA reconstructions of the spectra are overplotted in black. Note the failure of the reconstructions to follow the majority of the BAL features due to the iterative removal of these objects from the sample used to create the eigenspectra. Unusual quasars not used in this study are identified by the large χ^2 between the true and reconstructed spectra in wavelength regions around the C IV $\lambda1550$, Mg II $\lambda2796$ and Fe II $\lambda2600$ lines. The y-axis is in units of 10^{-17} erg cm^{-2} s^{-1} Å^{-1}. The SDSS spectroscopic filename (Modified Julian Date, plate number, fibre number) and quasar redshift are given above each panel.

quasar candidates using a simple "colour cut", based on the difference in the flux contained within bandpasses on either side of the C IV $\lambda1550$ emission line (or the Mg II $\lambda2796$ line when C IV is not covered); these obvious BAL quasars were then temporarily removed from the sample. The remaining spectra were shifted to the rest frame (without rebinning) and a PCA performed[8] on the spectra in each redshift bin using the "gappy-PCA" method of Connolly & Szalay (1999). Gappy-PCA is an iterative procedure in which missing pixels in each spectrum[9] are replaced by the reconstructed spectra created from the eigenspectra of the previous iteration. During reconstruction each pixel is weighted by the inverse of the squared error array, with missing pixels given zero weight.

A 41-pixel median filter was run through each quasar spectrum to identify narrow absorption features. The spectra were then reconstructed from the eigenspectra with these narrow features masked to prevent them from biasing the reconstructed continua low[10]. Reconstructed continua were created for the entire sample of quasars, including those excluded during the creation of the eigenspectra by the colour cuts around the C IV $\lambda1550$ and Mg II $\lambda2796$ lines. For those interested in the "BALnicity" of individual objects (e.g. Weymann et al. 1991), the quasar spectra are best reconstructed using only those eigenspectra visually identified as containing no BAL features, with the regions in which BAL features are expected masked during reconstruction so as not to bias the continua too low. Fig. A1 shows examples of reconstructions for both BAL and non-BAL quasar spectra.

The number of eigenspectra used in a reconstruction is not defined by the PCA procedure. For our purposes of identifying and removing unusual spectra, the precise number was not critical: BAL spectra and quasar spectra with associated absorption possess large χ^2 values between the true and reconstructed spectrum in the region of BAL features even when as many as 10-15 eigenspectra are used for reconstruction. This reflects the large variations in size and shape of BAL features which the PCA is unable to reconstruct even with relatively large numbers of components. As we were not interested in the precise classification of each object, we simply plotted the distribution of χ^2 between the true and reconstructed spectra in wavelength regions blueward of the C IV $\lambda1550$, Mg II $\lambda2796$ and/or Fe II $\lambda2600$ lines. All objects which fell into the high χ^2 tail of each distribution were removed, producing a new input sample for a second iteration of the PCA. The entire PCA and reconstruction procedure was repeated twice to ensure that as few

[8] Because of the large number of pixels in the SDSS spectra, we employed an expectation maximisation algorithm (Roweis 1997) which affords a substantial increase in speed.

[9] Pixels can be "missing" from a spectrum either because of defects or intrinsic rest frame wavelength coverage.

[10] Masks were applied during the gappy-PCA procedure by giving the pixels zero weight.

BALs as possible were present when creating the final eigenspectra and the eigenspectra were therefore most representative of ordinary quasars. In total, 13% of quasars, with redshifts $0.85 \leqslant z \leqslant 3.2$, were removed from the input sample.

Ultimately, our aim was to create a subsample of quasars suitable for: (a) searching for intervening absorption systems and (b) creating control spectra with which to compare the SEDs of quasars with absorbers. The final results reported in this paper are insensitive to the details of the scheme adopted to identify and exclude BAL quasars. By inadvertently removing any non-BAL quasars from the sample we would be reducing the sightlines available for the search, but this is a small effect given the size of the input sample. On the other hand, any BAL quasars which remain in the sample will increase the variance among the quasar SEDs, so reducing our ability to isolate any systematic differences between the SEDs of quasars with intervening metal absorption line systems and the quasar population as a whole. This effect is included in the Monte Carlo simulations used to estimate the mean and variance of SED colours for random samples of quasars (see Section 5.2).